Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Inferential Behavior in Children: I. The Influence of Reinforcement and Incentive Motivation: H. H. KENDLER, T. S. KENDLER, S. S. PLISKOFF, AND M. F. D'AMATO 207
Pre-Extinction in Sensory Preconditioning: W. J. COPPOCK 213
Context Effects With Judgmental Language That Is Absolute, Extensive, and Extra-Experimentally Anchored: D. T. CAMPBELL, N. A. LEWIS, AND W. A. HUNT 220
Time Sharing Between Frequency-Coded Auditory and Visual Channels: E. T. KLEMMER 229
A Correlational Study of Two Reasoning Problems: L. BRUNK, E. G. COLLISTER, C. SWIFT, AND S. STAYTON 236
Eyelid Conditioning as a Function of Unconditioned Stimulus Intensity and Intertrial Interval: W. F. PROKASY, JR., D. A. GRANT, AND N. A. MYERS 242
Verbal and Motor Responses to Seven Symbolic Visual Codes: A Study in S-R Compatibility: E. A. ALLUISI AND P. F. MULLER, JR. 247
"Permanent" Satiation Phenomena With Kinesthetic Figural Aftereffects: M. WERTHEIMER AND C. M. LEVENTHAL 255
Transfer to a Motor Task as Influenced by Conditions and Degree of Prior Discrimination Training: A. E. GOSS AND N. GREENFELD 258
Retention in Motor Learning as a Function of Amount of Practice and Rest: J. C. JAHNKE ... 270
Acquisition and Extinction of Problem-Solving Set: R. A. GARDNER AND W. N. RUNQUIST ... 274
Concept Formation as a Function of Competition Between Response Produced Cues: H. H. KENDLER AND A. D. KARASIK 278
Time-Order Errors in Comparative Judgments of Hurtfulness: R. H. GEERTSMA 284
The Influence of Complexity and Novelty in Visual Figures on Orienting Responses: D. E. BERLYNE 289
Interaction Between Total Stimulus Information and Specific Stimulus Information in Visual Recognition: J. R. NEWBROUGH 297
Supplementary Report: Effect of Addition of Irrelevant Verbal Cues on Perceptual-Motor Learning: H. W. BRAUN AND A. W. BENDIG 301
Erratum 302

American Psychological Association

Vol. 55 No. 3 March 1958

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Circulation Manager
SADIE J. DOYLE, Editorial Assistant

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor: Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices in this Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1958 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1958 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 55, NO. 3 MARCH, 1958

INFERENTIAL BEHAVIOR IN CHILDREN: I. THE INFLUENCE OF REINFORCEMENT AND INCENTIVE MOTIVATION [1]

HOWARD H. KENDLER
New York University

TRACY S. KENDLER
Barnard College

STANLEY S. PLISKOFF AND MAY F. D'AMATO
New York University

It has been reported (**2**) that preschool children are capable of inferential behavior. This conclusion was drawn from the results of an experiment in which *S*s received three separate experiences in a maze-like situation. Two of these experiences (A-B and X-Y) led to subgoals while the third experience (B-G) led to the major goal. The subgoal in the A-B experience served as the start of the B-G experience. On the test trial *S* was given a choice between A and X with the instructions to obtain G. The selection of A was considered the correct test response since A led to B which in turn led to G. A significant percentage of *S*s chose A on the test trial, thus suggesting that they were able to link the two separate habits based on the A-B and B-G experiences.

The present study was designed to determine whether inferential behavior resulting from habit linkage is influenced by two of the basic variables that govern behavior based on simple habits, e.g., bar pressing in rats. These variables are motivation and reinforcement.

It has been demonstrated that learning in relatively simple situations is dependent on some reinforcement mechanism but that such learning will be manifest only under appropriate motivating conditions. For example, in behavior based upon the simple habit of bar pressing, it has been shown that performance is a joint function of past training and present motivation (**4**), i.e., habit and drive variables. Animals with

[1] This research was sponsored by a grant from the National Science Foundation. The two senior authors are responsible for the design of the experiments and the preparation of this report. They are indebted to the junior authors for their aid in conducting the experiment and offering valuable suggestions. We are all indebted to Eugene Serson, who designed the inference apparatus. Acknowledgment is also made to the following nursery schools in Nassau County and New York City for their kind cooperation in supplying *S*s and facilities for the execution of this research: Rolling Hills, Shibley, Les Coquelicots, Parkside, Highbridge, Hoffman, Peter Pan, Riverdale Play School, Riverdale Neighborhood, Sedgewick.

limited training (five reinforcements) but high motivation (22-hr. deprivation) exhibited definite bar pressing response tendencies, while those with the same amount of training and low motivation (3-hr. deprivation) responded at practically a zero level.

If the learning-performance distinction (**5**) operates in inferential behavior of preschool children, then it would be expected that the independent variation of both reinforcement variables during training and motivational variables at the time of test should affect the amount of inference demonstrated.

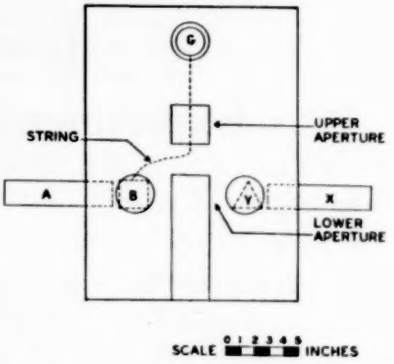

FIG. 1. Basic plan of inference apparatus.

METHOD

Subjects.—A total of 384 preschool children whose ages ranged from 36–60 months served as *S*s for the three experiments reported. They were recruited from nursery schools throughout the New York metropolitan area. They were randomly assigned without regard for their age or sex.

Apparatus.—The apparatus was similar to the one used in the initial study in that it lent itself to the same sort of procedure, i.e., a sequence of three separate experiences (A-B, X-Y, B-G) followed by a test trial involving a choice between A and X with instructions to the *S* to get G. The present apparatus differed markedly in terms of its structure and design. It was considered superior to the initial apparatus (**2**) because of its greater flexibility and ease of manipulation.

The basic plan of the apparatus is presented in Fig. 1. The apparatus is a gray box-like structure with a lever projecting from each of its vertical sides. The face of the apparatus had two rectangular apertures and three transparent windows that could be covered. The levers, painted black or white, when pushed inward, moved small boxes (either a yellow square box or a red triangular box) along a wooden track into the lower aperture. One of these boxes was attached by a hidden string to a translucent circular plastic container mounted on wheels. When this box was pulled down, the plastic container was drawn into the upper aperture.

The levers are referred to as A or X. The box pushed by A is referred to as B, while the one propelled by X is known as Y. The B box is the one attached to the circular plastic container known as G.

Experimental design and procedure.—During training the three round windows were covered. Each *S* received three different experiences. Initially *S* received two X-Y experiences with a raisin in Y on each trial. Then two A-B experiences followed with a raisin in B on each trial. Finally *S* received two B-G experiences.

Two major variables were used in these experiments: reinforcement during training and motivation during the test trial. Both of these variables were manipulated in relationship to the B-G experience. There were two reinforcement conditions. Positive reinforcement (R_+) occurred when *S* pulled B and G appeared in the lower aperture containing a gold-colored charm (purchased from Folz Vending Co., 990 Long Beach Road, Oceanside, N. Y.). Zero reinforcement (R_0) existed when G was empty during the B-G experiences.

The second experimental variable had to do with the motivational conditions at the time of the test. During the test trial when *S* was instructed to push A or X in order to get G, the cover over G was removed so that *S* could perceive what, if anything, was in G before it was moved down the track and thus made available to *S*. A positive incentive-motivation condition (M_+) existed if G contained a charm. If G was empty, a zero incentive-motivation condition (M_0) prevailed.

Three separate experiments were conducted. Experiment II replicated Exp. I, while Exp. III involved a slight modification designed to obtain a more sensitive measure of the effect of the experimental variables.

In Exp. I and II a 2×2 factorial design (R_+M_+, R_+M_0, R_0M_+, R_0M_0) was used. Each of the four groups contained 32 *S*s, resulting in a total of 128 *S*s for each experiment.

Experiment III contained only two groups. Group R_+M_+ during Exp. III received the same treatment as did Group R_+M_+ during the initial two studies. The *Ss* in Group $\sim R_+M_+$ were randomly assigned to one of the following procedures: R_+M_0, R_0M_+, or R_0M_0. Experiment III compared *Ss* that were under the combined influence of both positive reinforcement and motivational conditions with those who were not. There were 64 *Ss* in each of these two groups.

The experimental sessions were conducted in the various nursery schools from which *Ss* were drawn. While the external conditions varied from school to school, the *S* and *E* were always alone in a secluded room during the experiment. When *S* entered the room the apparatus was covered with a gray curtain. The *E* said, "We are going to play a little game. In this game if you do what I tell you, you will have a good time. Ready?" When *S* agreed, *E* removed the curtain and said, "Now [*S*'s name], do you see this stick [*E* points to X]? Push it in and see what happens. See? A raisin! You can have it, and you can eat it if you want to or save it in that [*E* points to a small paper cup]. Now do that again." In this second X-Y trial and any subsequent A-B trial *E* said, if the child failed to push the stick, "Go on, push the stick." If the child did not take the raisin when it became available *E* suggested, "You may have it."

The *S* then received two trials on the A-B sequence. The *E* pointed to A while saying, "All right, now push this one. See? Another raisin! Now do that again."

The B-G experiences were given with the following instructions: "Now look in here [*E* points to the square opening from which G is later obtained]. Now pull this cup [pointing to B] to you and see what happens. That's the way. Look at what's in here now [E points]." The *E* told the R_+ *Ss*, "You can have it." The R_0 *Ss* were told to touch G so that both R_+ and R_0 *Ss* would have essentially the same amount of contact with G. On the second B-G trial *S* was told, "Now look here again. Okay, pull the cup. Look." If *S* did not take the charm (R_+) or touch G (R_0), he was told to do so.

Immediately after the completion of the second B-G experience *S* received the test trial. On the test trial *E* said, "[Name], I want you to see what's here [*E* points to the round glass window below which the G cup is located]. See? Now listen to me very carefully. Don't do anything until I tell you to. Which *one* of these two sticks [*E* points to sticks with both hands simultaneously] must you push to get this [*E* points to window with both hands]? Remember, you may push only one. All right, now go ahead." The *E* was careful to see that *S* looked at both sticks. If the child's gaze remained focused on one stick *E* repeated, "See? There are two sticks."

If *S* pushed a stick after he received the test instructions *E* remained quiet for a period of about 10 sec. If *S* did not make a response during this period *E* removed the stick that had not been pushed and said "What must you do to get this [pointing to G]?" If there was still no response, *E* waited another 5 sec. and repeated the instructions.

Results

Since the performance of each of the four experimental groups did not significantly vary from Exp. I to Exp. II, the results of these two studies are combined and presented in Table 1. Four different response measures are reported. An A response refers to the selection of the A stick on the test trial. I_s is an abbreviation for a *spontaneous* inference response. An I_s response involved *S* pushing the A stick and then spontaneously pulling B to obtain G. An I_p response refers to the behavior of an *S* who pushes A and then waits until *prodded* by *E* with the question "And now what do you do?" before pulling B. The I_s and I_p responses therefore represent integrated inferential behavior as compared to an A response which merely indicates the initial correct response in an inferential sequence. The $\frac{I_s + I_p}{A}$ ratio represents the proportion of *Ss* who, after making the initial correct choice, completed the inferential chain.

The data reported in Table 1 lends itself to a number of possible comparisons. Starting with the reward variable, the data show that motivation without reward (R_0M_+) was inferior to motivation with reward (R_+M_+) for all four response measures. However, only the difference

TABLE 1

PERCENTAGE OF VARIOUS TEST RESPONSES FOR COMBINED GROUPS IN EXP. I AND II

(64 *Ss* in Each Group)

Group	Test Responses			
	A %	I_s %	$I_s + I_p$ %	$\frac{I_s+I_p}{A}$ %
R_+M_+	67	27	38	56
R_+M_o	52	13	17	33
R_oM_+	58	16	20	35
R_oM_o	58	16	25	43

involving the $I_s + I_p$ response measure was significant ($P < .02$) in terms of a two-tail test. Considering the motivational variable, reward without motivation (R_+M_o) was inferior to reward with motivation (R_+M_+). In this case the differences involving the I_s, $I_s + I_p$, and $\frac{I_s + I_p}{A}$ response measures were statistically significant at the .05, .01, and .05 levels, respectively. All of the test response measures of Group R_+M_+ exceeded those of Group R_oM_o but none of the differences was significant at the .05 level. None of the differences involving Groups R_+M_o, R_oM_+, and R_oM_o approached statistical significance. The only group that during the test trial chose the A stick at a level significantly greater than chance was Group R_+M_+ ($P < .01$).

The combined results of Exp. I and II suggested that (*a*) the optimal conditions for inference behavior involved both positive reinforcement and positive motivational conditions, and (*b*) the test performances of *Ss* lacking one or both of these conditions were essentially similar. It was therefore decided to do Exp. III, in which Group R_+M_+ was compared with the other three groups combined (Group $\sim R_+M_+$). The prediction, consistent both with the theoretical distinction between learning and performance as well as the suggestive findings of Exp. I and II, was that Group R_+M_+ would demonstrate test performance superior to Group $\sim R_+M_+$. Table 2 reports the results of Exp. III. The direction of results are consistent with the expectations, viz., children under the joint influence of both positive reinforcement and motivational conditions would demonstrate superior test performance. The only difference that was significant was the one involving the $I_s + I_p$% ($P = .04$).

Table 3 reports the combined data from all three experiments. In this table *Ss* are divided into two groups: Group R_+M_+ and Group $\sim R_+M_+$. This means that for the first two experiments *Ss* in Group R_+M_o, R_oM_+, and R_oM_o were combined into a $\sim R_+M_+$ group having three times as many *Ss* as Group R_+M_+. It will be recalled that in both Exp. I and II no significant difference existed in the performance of these three groups. In addition to this it was noted that the results from experiment to experiment were remarkably stable. In not one case did the performance of either group differ significantly from one experiment to another, nor is

TABLE 2

PERCENTAGE OF VARIOUS TEST RESPONSES IN EXP. III

(64 *Ss* in Each Group)

Group	Test Responses			
	A %	I_s %	$I_s + I_p$ %	$\frac{I_s+I_p}{A}$ %
R_+M_+	63	20	34	55
$\sim R_+M_+$	55	9	18	34

TABLE 3

PERCENTAGE OF VARIOUS TEST RESPONSES FOR COMBINED R_+M_+ AND $\sim R_+M_+$ GROUPS IN EXP. I, II, AND III

Group	N	Test Responses			
		A %	I_a %	$I_a + I_p$ %	$\frac{I_a + I_p}{A}$ %
R_+M_+	128	65	23	36	55
$\sim R_+M_+$	256	56	13	20	37

there any overlap between the scores of Groups R_+M_+ and $\sim R_+M_+$ for any response measure. Although this *ad hoc* combination of data is open to criticism, it was felt that it was justified for pragmatic reasons. The present experimental technique is rather novel and the results that it produces should be subjected to a penetrating scrutiny to uncover trends that might be hidden by more gross analyses.

Using a two-tail test to evaluate the reliability of the differences, it was found that significant differences were obtained between Groups R_+M_+ and $\sim R_+M_+$ with the I_a ($P = .02$), $I_a + I_p$ ($P < .01$), and $\frac{I_a + I_p}{A}$ ($P < .01$) response measures. The difference between the two groups in terms of the A response was not significant ($P = .08$). Of particular importance is the finding that the proportion of A responses in *both* groups was significantly greater than chance (at the .05 level for Group $\sim R_+M_+$ and at the .01 level for Group R_+M_+).

DISCUSSION

This study supports the idea that the theoretical distinction between learning and performance has its counterpart in the inferential behavior of children. A child who is capable of demonstrating appropriate inferential responses may not perform them because of inadequate motivational conditions. Presumably *S*s in Group R_+M_0 learned as much about the two habits (A-B and B-G) of the inferential sequence as did *S*s in Group R_+M_+, but because of the absence of the positive incentive motivational condition at G, Group R_+M_0 demonstrated inferior inferential performance on the test trial.

In the simple bar-pressing behavior of rats it is assumed that the hunger drive activates habit to produce performance (**1**). If either drive or habit is weak, so is performance (**3, 4**). Our present results cannot be so easily interpreted. The operation of placing a charm at G and permitting *S* to see it before making his test response can influence his inferential performance in two major ways: the M_+ condition can (*a*) increase the tendency to initiate the inference sequence by raising the probability of *S* selecting A; (*b*) increase the tendency to integrate the two habits by raising the probability of *S* pulling B when it comes into view after pushing A. The fact that the R_+M_+ group was significantly superior to Group R_+M_0 in terms of the $\frac{I_a + I_p}{A}$ response measure suggests that the M_+ condition operates in the second way. The fact that Group R_+M_+ was not significantly superior to Group R_+M_0, or for that matter any of the other groups in the selection of the A response suggests that it does not operate to increase the probability of selecting A. There are some reservations to the acceptance of this latter conclusion. It may be that the differences between Group R_+M_+ and the other groups in making the A response were attenuated by chance factors. Since the *S* is instructed to select the A or X stick, the A-response measure is more susceptible to the action of chance factors than are the other response measures. It may well be that the M_+ condition both initiates and integrates inferential behavior. A more definitive answer will have to await a more definitive test.

There is another important difference

between the present results and those obtained with animals in an instrumental conditioning situation (3, 4). In the instrumental conditioning situation if habit is approximately zero or drive is approximately zero, then performance becomes approximately zero (defined in terms of the number of bar pressings to reach extinction). Such was not the case in the present experiment. The absence of a positive habit condition (R_o) or a positive motivational condition (M_o), or both, did not reduce inferential behavior to zero. The combined Group $\sim R_+M_+$ performed the A response significantly greater than chance. Twenty per cent of these *S*s were able to demonstrate an integrated sequence of inferential responses.

One may conclude from such evidence that inferential behavior can occur in the absence of positive reinforcement or motivational conditions. There is, however, a more reasonable alternative. Obviously reinforcement and motivational systems of preschool children are complex. It would be naive to think that tokens we used represented the only reinforcements and were the basis of the child's only drive system. The *S*'s motivation to please the adult *E*, to follow instructions, and the reinforcements derived from these motivations can be expected, for the present, to play a significant role in research of this sort. It was consequently not our intent to provide an all-or-none type of experimental control of the reinforcement and motivational variables. Instead we sought merely to discover whether inferential behavior, like the simple bar-pressing habit in animals, is influenced by independent variations of reinforcement and motivational variables. This question was answered in the affirmative and it was found that the optimal condition for the performance of inferential behavior involved the joint action of reinforcement and motivational variables.

Summary

This study was designed to determine whether inferential behavior resulting from habit linkage is influenced by reinforcement and motivational variables. Three studies, involving 384 preschool children, were executed in an experimental situation that provided them with three separate experiences. Two of these experiences (A-B and X-Y) led to subgoals (raisins in distinctive boxes), while the third experience (B-G) led to a blue container. The subgoal (B) in the A-B experience served as the start of the B-G experience. During training, G (the blue container) held the major goal (an attractive charm) for some *S*s while it was empty for other *S*s.

On the test trial *S* was confronted with a choice between A and X. Choosing A would lead to B which if pulled would produce G. During the test trial the blue container (G) could be seen through an opening in the apparatus. For some *S*s it held the major goal while for other *S*s it was empty.

The results indicate that *S*s who received the major goal at G during the learning trials *and* who could see the major goal at G during the test trial performed better than *S*s for whom only one or neither of these conditions prevailed. It was concluded that inferential behavior, like the simple bar-pressing habit in animals, is influenced by independent variations of reinforcement and motivational variables.

References

1. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
2. Kendler, H. H., & Kendler, T. S. Inferential behavior in preschool children. *J. exp. Psychol.*, 1956, **51**, 311–314.
3. Koch, S., & Daniels, W. J. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1945, **35**, 167–187.
4. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.
5. Tolman, E. C. *Purposive behavior in animals and men.* New York: D. Appleton-Century, 1932.

(Received April 8, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

PRE-EXTINCTION IN SENSORY PRECONDITIONING[1]

W. J. COPPOCK[2]
University of California

Sensory preconditioning experiments usually consist of three phases. In the first, called the "preconditioning" phase, *S* is repeatedly presented with a pair of stimuli, for example a faint tone and a weak light. In the second, called the "training" phase, one of these two stimuli, the light, is repeatedly paired with an unconditioned stimulus such as a shock until a stable conditioned response—for example, a GSR—has been established. In the third, or "test" phase, the other stimulus from the preconditioning phase, the tone, is presented alone. Figure 1 summarizes this procedure in the top row labeled PC group. Brogden (**1**), the first to perform this kind of experiment, found his dogs making the CR acquired during the training phase when presented with the test stimulus. An appropriate control group was run in which the tone and light were repeatedly presented but never paired. Figure 1 summarizes this under procedure for the C group. Parentheses enclose the unpaired tone (T) and light (L) to distinguish them from the paired stimuli of the PC group. Failure of Brogden's animals in such a control group to give significant responses in the test phase demonstrated that the PC group's test phase performance was dependent upon some kinds of associations formed between the paired stimuli presented in the preconditioning phase. The central question is, then: What kinds of associations are formed in the preconditioning phase and how do they serve to mediate the response in the test phase?

An account of the way the test stimulus comes to evoke a response, apparently without ever being conditioned to it, has been given by Osgood (**5**). This explanation, which will be called the "SR Conditioning" hypotheses, ascribes preconditioning to processes familiar in classical conditioning.[3] Figure 1 diagrams this SR Conditioning account of the performance of the various groups we shall consider. According to this view each of the two stimuli in the preconditioning phase of the PC group has a distinct UCR with an attendant response-produced stimulus. For example, each time *S* sees the light (S_2) his pupil contracts (R_2) and this response produces distinct sensory feedback (s_2). During the preconditioning phase the tone (S_1) becomes conditioned to pupil contraction (R_2). In the training phase the light is presented, the pupil contracts, and stimuli from that response (s_2) also become conditioned to the measured response, R_2. In the test phase the tone elicits pupil contraction and the stimuli from this response, s_2, elicit R_2.

Thus the SR Conditioning hypothesis assumes that CR's are formed in the preconditioning phase. If these hypothetical CR's are like all other CR's, then certain standard experimental procedures should weaken their strength. Accordingly, if the preconditioning phase is followed by a new phase, called the

[1] This paper is based on a Ph.D. dissertation presented to the University of California. The author wishes to express his deep appreciation to D. A. Riley, B. F. Ritchie, and L. E. Davis for their generous help throughout the project.

[2] Now at University of Oregon Medical School.

[3] Osgood does not make a distinction between two possible SR positions, the one here called the SR Conditioning hypothesis, and the Wickens and Briggs account (**6**) which can be called an SR Common Response interpretation.

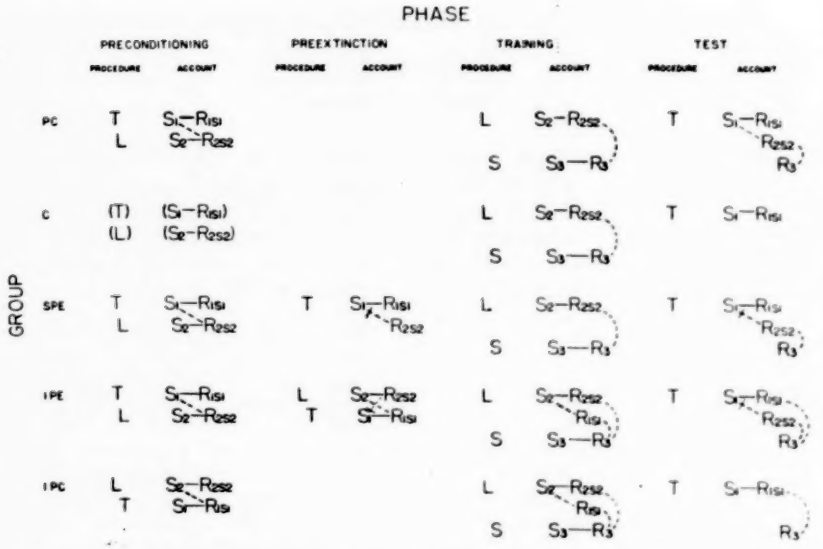

FIG. 1. Diagram of the SR Conditioning hypothesis. L, light; T, tone; S, shock; —, unconditioned bond; - - -, conditioned bond; + - -, partially extinguished bond; ( ), non-paired stimulus; displacement of paired stimuli on horizontal axis indicates delay.

"pre-extinction" phase, in which the first stimulus is presented but the second no longer follows it, then whatever CR is associated with the first stimulus should be weakened more and more as such presentations are repeated. There are two ways in which such a pre-extinction procedure can be carried out. First, as is shown by the SPE group in Fig. 1, the tone may be presented alone in the pre-extinction phase. This procedure will be called "Simple Pre-extinction." Second, as is shown by the IPE group, the tone may be presented following the light. (Note that horizontal displacement of symbols is used to indicate a delay between presentation of stimuli.) This procedure will be called "Inverted Pre-extinction." In both cases the S_1-R_2 association should be weakened. As a consequence, the more this association is weakened, the less one should observe the mediated response, R_3, in the test phase. But there is one consequence of the treatment of the IPE group which should counteract this. In that group during the pre-extinction phase, S_2 is followed by S_1 and therefore S_2 will come to elicit R_1. Then in the training phase in which S_2 is the CS for the new response R_3 as well as for R_1, the response-produced stimulus, s_1, will also become conditioned to R_3. Generally the training phase quickly establishes a strong association between S_2 and R_3. Efficient learning of this SR association suggests, in the absence of additional information, that conditions would be suitable for efficient learning of the s_1-R_3 association. Therefore, in the test phase for the IPE group there is an additional mediator, R_1s_1, for the test response and a mediator of maximal strength. Clearly this hypothesis predicts that in the test phase S_1 will elicit R_3 to a greater degree in the IPE group than in the SPE group.

One other group is summarized in Fig. 1. In this IPC group two stimuli are also presented in the preconditioning phase. However, S_2, the stimulus presented first in the preconditioning pairings, is conditioned to R_3 in the training

phase. S_1, the stimulus presented second, appears in the test phase. This order of stimulus presentation in the preconditioning phase is the inverse of the order used in the PC group, hence this is called the "Inverted PC" group. Now, according to the SR Conditioning hypothesis, the preconditioning phase pairings in this group condition S_2 to R_1. Then in the training phase S_2 elicits R_1 with the result that s_1 becomes conditioned to R_2. In the test phase R_2 is elicited by the mediator R_1s_1. There are two reasons for thinking that this mediation through R_1s_1 should be strong. As mentioned above, the training phase generally is suitable for efficient learning of SR associations. Also, note that in the test phase the mediator for the IPC group, R_1s_1, is elicited as a UCR of the test stimulus whereas the mediator for the PC group, R_2s_2, is a CR of the test stimulus. This substitution of an unconditioned for a conditioned association should enhance the test phase performance of the IPC group. Thus mediation through R_1s_1 in the IPC group should be strong, probably at least as strong as mediation through R_2s_2 in the PC group.

The IPC group is purposely not called a backward preconditioning group. The term backward preconditioning is not appropriate since this analysis has shown that test phase responses in this group can be due to associations formed by a CS preceding a UCS—that is, forward conditioning.

Four predictions have been developed about the relative strength of test phase responses in these five groups: (*a*) IPE greater than SPE since the IPE group has an additional mediator, R_1s_1, for the test response; (*b*) SPE equal to or greater than the C group since the partially extinguished association S_1-R_2 gives the SPE group a possible mediator R_2s_2; (*c*) IPC greater than C group since the IPC group has the mediator R_1s_1 for the test response; and (*d*) PC greater than C group since the PC group has the mediator R_2s_2.

Method

Subjects.—The *S*s were 95 students from the elementary psychology course. Nineteen *S*s were assigned to each of the five experimental groups.

Apparatus.—Tones of 1024 cps generated by a standard audio oscillator were fed to a speaker in front of *S*. The intensity was approximately 40 db above .0002 μ bars. The light, a 7½-w. frosted bulb near the speaker, was judged by *E* to be about as salient as the tone. The shocking apparatus could be adjusted to deliver up to a 3-ma., 60-cps shock. Durations of these stimuli and the interstimulus delay were controlled by a two-channel decade timer.

The palm and back of each of *S*'s hands were rubbed with EKG paste. Zinc electrodes (1-in. diameter) with attached salt-saturated sponges were fastened over the rubbed areas. Shock was administered to the left hand. A constant current of 20 μamp. was fed to the electrodes on *S*'s right hand (**2**). Variations in voltage-drop across *S* resulting from changes in skin resistance were fed to a direct coupled amplifier. The output of the amplifier drove an Esterline-Angus recording milliammeter which was calibrated in terms of *S*'s resistance. The equipment could be balanced over the range 0 to .1 megohm basal skin resistance. Five sensitivities were available with a maximum of 410 ohms change per inch deflection on the recording chart. The chart was run at a speed of 5 sec. per inch travel.

Procedure.—The *S* was seated in a soundproofed room which was indirectly lighted by a 75-w. bulb. The *S* was told that some electrodes were being attached to his hands to record "physiological responses." He was warned that there would be a mild tingling sensation on his left hand now and then, and told that he should signal *E* if it became too painful, in order that the experiment might be discontinued. No *S* so signaled. The final instructions were:

"This is what I want you to do. Make yourself comfortable so that you can sit quietly for the half hour or so the experiment takes. From time to time this light will go on. From time to time you will hear a tone from this speaker. From time to time you will feel the prickling sensation on your left hand. Try to stay as alert and attentive as you can and watch for the light, listen for the tone, and be aware when the sensation occurs on your left hand. All set?" (If *S* had any questions they were answered. Then the doors were closed.)

At this point *S* was assigned to a group by a random process. This assignment, made after *S* was in the soundproofed room, eliminated the possibility of introducing bias during *E*'s interaction with *S*.

Three minutes after *S* was isolated, the first stimulus was given. The PC group had T and L paired 10 times in the preconditioning phase. T was on for 3 sec. L overlapped the last 2 sec. of T. The intertrial interval was randomized around 30 sec. Two minutes after the last preconditioning trial, L was paired with S in classical conditioning. L was on for 2 sec.; S was on during the last second of L. The intertrial interval was randomized around 1 min. The intensity of the first shock was 1½ ma., the second, 2 ma., the third through tenth and last, 2½ ma. Two minutes after the last training trial the first test trial was given in which T was presented by itself for 3 sec. Five such test trials were given at a random intertrial interval of 1 min.

The C group had L and T presented 10 times each in a prearranged staggered order. T was on for 3 sec., L for 2. Training and test phases for this group are identical with these phases in all other groups.

The procedures for the SPE and IPE groups differ from that for the PC group only in the inclusion of a pre-extinction phase. In the SPE group T was presented for 3 sec. by itself 2 min. after the last preconditioning trial. Ten such trials were given at a randomized intertrial interval of 30 sec. In the IPE group T and L were paired with L on for 3 sec. and T overlapping the last 2 sec. of L.

Finally, the IPC group differs from the PC group only in that, in its preconditioning phase, T followed the onset of L, not vice versa.

In order to avoid confusion in Fig. 1 and the preceding discussion, the counterbalancing of each of these groups has not yet been mentioned. This counterbalancing was accomplished by using the tone for S_1 and the light for S_2 in part of each group, and the light for S_1 and tone for S_2 in the rest of each group.

Scoring.—Records of palmar-dorsal skin resistance were made during each phase. The magnitude of the response was calculated from the basal resistance at the time of stimulation (B) and the minimal resistance reached during the next 15 sec. (M). These values were read to the nearest 100 ohms during the training phase and to the nearest 10 ohms where possible during the test phase. Results are expressed in terms of percentage change in skin resistance, $\frac{B - M}{B} \times 100$.

Results

Marked positive skewing was evident in the scores during the training and test phases. Accordingly, distribution-free tests were used exclusively in the analysis of these data. Significance levels for the tests were calculated by making a normal approximation based on the published formulas for the expectation and variance of the test statistic.

A preliminary test for the effect of variation in test stimulus (i.e., T vs. L) on response magnitude was made by ranking methods for groups of replicates (7). The smaller obtained sum of ranks of 407 has a probability of a chance occurrence of .4. Since there is thus no reason to believe that the particular choice of stimuli used affected the results, no further reference will be made to the counterbalanced subgroups.

Figure 2 shows the median response for each group for each of the five test trials. Group SPE gave a smaller median response than Group IPE on each trial. Group PC overlapped both the SPE and IPE groups. Group C gave a considerable median response on the first test trial but little or no response thereafter. Group IPC performed only slightly

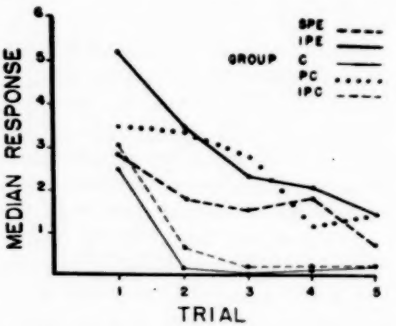

FIG. 2. Median percentage GSR on the five test-phase trials.

TABLE 1

SUMMARY OF PAIRED REPLICATE TESTS OF TOTAL TEST-PHASE RESPONSE

Contrast	Sum of Negative Ranks	p
IPE-SPE	−55	.05
SPE-C	−49	.03
IPC-C	−78	.25
PC-C	−19	.001

better than Group C on the first four trials and precisely the same as Group C on the last trial.

Tests of paired replicates by ranks (**6**) were applied to the four pairs of groups called for by the predictions. For each test the differences between the sum of the five test-phase scores of an *S* in one group and a matched *S* in another group were ranked and summed. The matching variable was the sum of the last two training-phase responses. Negative signs were attached to negative differences, i.e., cases where, contrary to prediction, the second *S* out-performed the first. The sums of negative ranks and associated one-tailed probabilities are presented in Table 1.

The IPE-SPE contrast calls for acceptance at the 5% level of the alternative that inverted pre-extinction is superior to simple pre-extinction in maintaining test-phase performance. The SPE-C contrast tests for residual mediation through R_2s_2. The SPE group is significantly superior; hence, by implication, simple pre-extinction left some strength in the S_1-R_2 association. The IPC-C contrast tests for the predicted mediation through R_1s_1. The groups are not different; hence, mediation through R_1s_1 did not occur. Finally, the probability of .001 attached to the PC-C contrast indicates that preconditioning phase pairings significantly increased the magnitude of R_2 to S_1 in the test phase.

Three further tests were made on these data. The first tested for a change in magnitude of test-phase responses as succeeding *S*s were run. The test showed that on the average there was no change in response magnitude during the execution of the experiment. The second tested for interaction between response magnitude in the groups and the period at which the data were collected. The obtained χ^2_r (**3**) of 5.7 with 4 *df* has a chance probability of .2. Apparently there were no systematic effects between groups over time. Finally, a test for equivalence of groups on GSR reactivity during the training phase was made with a one-way analysis of variance by ranks (**4**). The obtained *H* of 3.35 with 4 *df* has a probability of .5 for a chance occurrence. There is little reason to suppose that the five groups were not sampled from the same population with respect to GSR reactivity.

DISCUSSION

Two of the results, when considered in light of the highly significant performance of the PC group, are especially crucial to the SR Conditioning hypothesis. The first, the predicted performance of the SPE vs. IPE groups, supports the hypothesis. The second, the failure of the IPC group to show any evidence of mediation in the test phase, points to an inadequacy of the hypothesis to account for all the sensory preconditioning phenomena. Both results require further comment.

Considerable overlap exists between the distributions of scores for Groups SPE and IPE. The analysis suggested that mediation through R_1s_1 would be strong enough to produce an obvious difference between the distributions. There are three factors which might have reduced the difference between groups. Two of them can be ruled out but the

third cannot. First, it could be argued that no SR associations had been formed in the preconditioning or pre-extinction phases. This is not the case, as is clearly shown by the significant difference between the test-phase performance of Groups PC and C. From this result the hypothesis must assert that a strong association was formed in the preconditioning phase. Second, the difference between Groups SPE and IPE might be small because of competition between light and tone responses. For instance, *S* might not turn his head toward the speaker (R_1) when the light (S_2) is on because for some reason he always fixates lights (R_2). In such a case stimuli from head turning (s_1) could not become conditioned to R_3 in the training phase. Consequently Groups SPE and IPE would perform similarly in the test phase because the only mediation possible would be through R_2s_2 which is of equal strength in both groups. Apparently, however, such competition did not occur, for a test between the counterbalanced parts of Group PC showed that it made no difference whether *S* was tested with the tone or light. Thus, R_3 occurred in the test phase regardless of whether R_1 was a response to light or tone. The hypothesis says that if Group PC outperforms Group C, R_2 must be occurring in the test phase. We can therefore conclude that R_2 occurred regardless of whether R_1 was a light or tone response, that is, that neither response dominated the other. Thus the smallness of the difference between Groups IPE and SPE cannot be due to domination of R_1 by R_2.

There is another way to account for the small difference between Groups SPE and IPE, and this reason may be the correct one. If the preconditioning phase resulted in great strength in the S_1-R_2 association, it is possible that the pre-extinction phase might weaken it only slightly. In this case the two groups would have strong mediation through R_2s_2 in the test phase. Additional mediation through R_1s_1 in Group IPE might then add only a small amount to the strength of R_3 to S_1. The available data lend some support to this interpretation. The performance of Group PC is quite superior to that of Group C, which, according to the SR hypothesis, implies considerable strength in R_2 to S_1. Also, the test between Groups SPE and C showed that the pre-extinction trials left significant strength in the S_1-R_2 association. This suggests that the procedure used developed a strong S_1-R_2 association which was not greatly aided by additional mediation through R_1s_1.

The results for Group IPC pose serious problems for the SR Conditioning hypothesis. This group performed no differently from Group C. The hypothesis predicted superior performance for this group and suggested reasons why it should perform as well as Group PC. Now it is apparent that this failure of IPC to perform cannot be due to lack of learning SR associations in the PC phase. It is also apparent that IPC's poor performance cannot be due to competition between R_1 and R_3. These arguments were presented above and need not be repeated here. Thus, according to the hypothesis, R_1 must occur in the training phase. But note (Fig. 1) that in the IPC group, R_1, the crucial mediating response, is elicited in the training phase as a CR to S_2. In Group PC, however, the mediating response, R_2, is elicited in the training phase as a UCR to S_2. Therefore, in Group IPC the association S_2-R_1 may weaken in the training phase so that R_1 occurs less and less frequently and s_1 has less than the full 10 pairings to strengthen its association with R_3.

There are two processes that might be responsible for such weakening. The first, internal inhibition, seems a less likely explanation than the second, external inhibition. Consider the first process—internal inhibition. In the IPC group's training phase, R_1 to S_2 might weaken from experimental extinction since the UCS for this association is not given. But if extinction occurs in the training phase, the pre-extinction phase should rapidly extinguish the S_1-R_2 association in Group SPE. But

the finding is that simple pre-extinction experience leaves considerable strength in R_3 to S_1, hence, according to the hypothesis, R_2 to S_1. It is difficult to see why SR associations should be hard to extinguish in the pre-extinction phase but easy to extinguish in the training phase, especially since the latter association is being practiced under conditions suitable for strengthening SR associations. Extinction during the training phases as an account of IPC's performance cannot solve the problem.

But there is a second reason why R_1 to S_2 may be weakened during the training phase. Even a strong SR association can be disrupted if a relatively intense stimulus is presented about the time the CS occurs. Perhaps the S_2-R_1 association is being externally inhibited by S_3. It is obvious that if this were happening the IPC and C groups would perform alike. Further work will need to be done before the legitimacy of this account can be assessed.

It is evident that Group IPC creates a problem for the SR Conditioning hypothesis. Reasons for the poor performance have been suggested, but lacking further evidence we must conclude that the IPC group casts doubt on the SR Conditioning account of sensory preconditioning.

Summary

The SR Conditioning hypothesis says that classical conditioning principles will account for sensory preconditioning. This interpretation was developed and four predictions obtained. Five groups of 19 human *Ss* were run. After preconditioning pairings of tone followed by light, two groups were "pre-extinguished" by presentations of tone alone, or pairings of light followed by tone. Two other groups were preconditioned with pairings of tone followed by light or light followed by tone, respectively, but had no pre-extinction experience. A control group had tone and light presented but never paired. After these presentations, each group was conditioned to give the GSR to light and tested for GSR's to tone. Three predictions were confirmed at the 5% level, but the group that was preconditioned with the inverted order of light followed by tone, was not different from the control group. This result seriously questions the ability of the SR Conditioning hypothesis to account for sensory preconditioning. Possible reasons why this result was obtained are discussed.

References

1. Brogden, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25,** 323–332.
2. Flanders, N. A. A circuit for the continuous measurement of the galvanic skin response. *Amer. J. Psychol.*, 1953, **66,** 295–299.
3. Friedman, M. A comparison of alternative tests of significance for the problem of m rankings. *Ann. math. Statist.*, 1940, **11,** 86–92.
4. Kruskal, W. H., & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. statis. Ass.*, 1952, **47,** 583–621.
5. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
6. Wickens, D. D., & Briggs, G. E. Mediated stimulus generalization as a factor in sensory pre-conditioning. *J. exp. Psychol.*, 1951, **42,** 197–200.
7. Wilcoxon, F. Probability tables for individual comparisons by ranking methods. *Biometrics*, 1947, **3,** 119–122.

(Received December 18, 1956)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

CONTEXT EFFECTS WITH JUDGMENTAL LANGUAGE THAT IS ABSOLUTE, EXTENSIVE, AND EXTRA-EXPERIMENTALLY ANCHORED [1]

DONALD T. CAMPBELL, NAN A. LEWIS, AND WILLIAM A. HUNT

Northwestern University

The interpretation of the context effects found when using the method of single stimuli is ambiguous because the usual methodology unnecessarily confounds two separable lines of inquiry: distortions in identity judgments or "illusions" on the one hand, and semantic effects attributable to the learning of a novel and limited set of response categories on the other. It seems clear that many of the early studies were focused on the distortion of identity judgments of such a nature as to be independent of the mode of response allowed. Thus in the classic demonstrations of thermal adaptation *S* does not recognize as the same the water temperature which he previously called "cool" and now calls "warm." The effect would be found even were *S* to respond in terms of estimated degrees Fahrenheit. Similarly, when *S* at one time reports "sweet" to a .03 saline solution and "salty" in another context, he is unaware of the identity of the samples and would show the effect even if asked to estimate on both occasions the percentage saturation of the solutions. Such distortions in identity judgments, illusions, or shifts in the appearance of the stimulus independent of judgment language, are similarly involved in Helson's demonstration that the same gray paper under the same green illumination will appear green with one background and pink with another (**8**). Something of this traditional interest remains in many current studies employing the method of single stimuli.

Beginning at least with the studies of anchoring effects (**9, 14**), and characterizing many recent studies (e.g. **5, 12**), is a semantic interest, focused on such processes as the categorizing of stimuli in terms of a novel and restricted set of response categories, the nature of the enforced redefinition of categories when stimuli or anchors outside of the original context are introduced, or the efficiency of information transferal under various coding conditions. This interest, reinforced by modern communications theory, is a separate but equally valid and legitimate concern.

In the method of single stimuli as currently applied, the major effects summarized under the principle of adaptation-level can be interpreted either as distortions of identity judgments which are independent of the specific details of the response system allowed, or as semantic effects limited to the specific response system employed in the method of single stimuli. The following four features contribute to this ambiguity:

A situationally relative response language.—Terms like *heavy* and *light* are in their proper semantic usage situationally relative, i.e., they convey no absolute meaning apart from

[1] This study is part of a larger project subsidized by the Office of Naval Research under contract 7onr-450(11) with Northwestern University. The opinions expressed here are those of the individual authors and do not represent the opinions or policy of the Department of the Navy.

a specific comparative setting. We can speak of a heavy truck, a heavy suitcase, or a heavy fish line weight. In such usage, terms like *heavy* and *light* contrast with "absolute" terms like *one ounce*, *ten pounds* or *three kilos*, which, in dealing with the same attributes of physical objects, have become extricated from specific, immediate comparisons and are understood to be invariant attributes of the object, appropriately descriptive of it no matter what its setting.[2] With a situationally relative response language, the specific defining instances of a given term change with each context, and in complete semantic consistency, *S* may call a weight "heavy" at the present moment which he called "light" a moment ago, meanwhile clearly realizing it to be the same weight.

The E's novel, arbitrary response language.—Typically in the method of single stimuli, *S* is given a novel response language which he is to learn to use by a trial-and-error process. The presence of this situation is frequently attested to by such *S* comments as, "Now I think I'm getting to know *what you mean* by 'very, very, very heavy.'" If under this language-learning set the context of stimulation is shifted, *S* naturally decides that he has been wrong about his definition of the terms in *E*'s arbitrary language, and accordingly may shift his application of judgmental terms without any necessary distortion of identity judgments were they to have been asked for.

The restricted number of response categories.—A third potential source of purely semantic effects lies in the limited response language provided. If *S* has adjusted his usage of the given relativistic response categories to the initial range of presented stimuli and is then presented with stimuli far beyond that range, he is forced to redefine the judgmental categories or to extend, on his own initiative, the response language. Sometimes *E*s include in their reports such bootleg categories. More often they enforce the language limitations, and there results a redefinition of judgment categories which can be independent of any distortion of identity judgments. (See Bevan and Darby [1] for an illustration of this problem.)

S's self-instruction to communicate differentiation.—If *S* is asked to state, for example (**10**), whether notes are "high" or "low," he probably assumes a self-instruction to communicate to *E* those differentiations he discriminates among the stimuli. If the range of stimuli then shifts so that now all notes are in the region he initially called "high," he may feel inadequate in continuing to give this single report when there are differentiations among the current stimuli which he is not reporting. Related to this are *S*'s expectations that *E* wants *all* of the original response categories used, and that *E* will present a symmetrical range of stimuli.

All of these semantic factors could operate to create the types of context phenomena characteristically found. If there are identity distortions or illusions over and above these semantic redefinitions, it would seem possible to extricate them by employing a judgmental language that was (*a*) absolute, in the sense of referring to a situationally invariant aspect of the stimulus, (*b*) extra-experimentally anchored, so that it would not appear to be an arbitrary and transient linguistic convention for laboratory purposes only, and (*c*) unlimited, extensive, or extendible in terms of the number of categories. Such a language is provided by the measurement categories of science and everyday life, in terms of *ounces*, *inches*, *per cent saturate saline solution*, *degrees Fahrenheit*, etc. These response categories have been shown to be appropriate and efficient for psychophysical purposes in the method of single stimuli in studies focused upon the precision of discrimination and Weber's law (**2, 13**). They seem

[2] This is not to claim the possibility of absolute terms in any epistemological sense. The meaning of *ounces* and *pounds* or of *inches* and *feet*, etc., are of course relative to standard comparison objects. But the employment of these stable and remote comparison objects makes possible the labeling of a present object with descriptive terms which remain invariantly appropriate to it in all specific contexts. Such descriptive terms are here designated as *absolute*, in contradistinction to such comparative terms as *heavy* and *light*, which semantically may be appropriate descriptive terms for an object in one setting and not in another. The previous employment of the term *absolute* in connection with the use of relativistic response terms in the method of single stimuli can only be regretted as inappropriate and confusing from our present perspective. The term *absolute* has been used here in lieu of a better term, and, it is hoped, in consistency with popular linguistic connotations and dictionary usage (*vide* **2**).

eminently appropriate for single-stimuli studies of the effect of context upon the distortion of identity judgments. The present study attempts this for judgments of pitch, by invoking the piano keyboard as a judgmental language for report. The major effects under investigation are the contrast illusion and loss of refinement created by drastic shifts in the context of stimulation.

Method

Because the present argument depends upon *S*'s understanding of his task, it seems appropriate to introduce the reader to the experiment as *Ss* were introduced to it in being given the following instructions:

"This is an experiment in judging pitch. In a few minutes you will hear a tape recording consisting of a series of tones. It will be your task to make a judgment as to the pitch of each one as it is heard on the recorder. The diagrammatic sketch of a piano keyboard is designed to help you make your ratings or judgments. You will notice that each key is labeled and that a dark line divides the notes into octaves. To simplify the rating task for you, you are not to use the names of the notes when reporting your judgments. Instead, you will use the corresponding numbers which appear to the right of each key on the sketch. For example, suppose the first tone you hear sounds to you as if it is a little bit above 'middle *C*' (the middle note of a piano keyboard) and you guess that it is specifically 'the $E^{\flat}$ above middle *C*.' In this case you would simply write the number '43' in the first box because the coded number for the $E^{\flat}$ above middle *C* is 43. Immediately after each tone is heard, you are to write the number which corresponds to the note which you think was played. You should not necessarily expect all of the notes indicated on the diagram to be used during this experiment. Each tone will be heard for 1½ sec. and will be followed by 6 sec. of silence. During this time you will make your ratings. We have found that 6 sec. provides plenty of time for a rating to be made. It does not allow time for undue contemplating or for changing a rating once made, however. Therefore, you must plan your time accordingly. Do not allow yourself to fall behind and do not miss making a rating. If strict attention is paid, neither of these misfortunes will befall you.

"Although this task may at first sound difficult to those of you who know little about music, we have found from past experience that you do as well as some people with musical training. Therefore, do not allow yourself to become discouraged. The task is easier than it appears."

The tones were produced by an electrical audio-oscillator and acoustically recorded on tape. The nine tones used in this experiment (see Table 1) were based on the equal temperament scale (i.e., there was a uniform frequency interval between notes based on the 12th root of 2).

The diagrammatic keyboard was on a piece of paper some 15 × 2 in., and represented the 88 keys of the standard piano, with note letters (e.g., *C*, $C^{\sharp}$, *D*, etc.) octave designations, and note numbers from 1 to 88. Prior to beginning the task, *Ss* filled out a one-page questionnaire concerning their musical training and activities. The questionnaire and the general instructions created in *Ss* an objective-test orientation in which they were attempting to report accurately the "true" or "absolute" nature of the stimuli rather than making relative comparisons or describing private experiences. Although the numerical reporting form was novel to *Ss* and kept them focused on the diagrammatic sketch throughout the experimental sessions, the definition of terms was clearly anchored to a public and pre-experimental reference system, rather than being an arbitrary creation of *E*. As presented, the system did not represent an infinitely extendible response language. It was hoped, however, that the vocabulary provided would be extensive enough so that no *S* would exhaust his repertory of terms in the initial phase and hence not have ample terms left to indicate more extreme stimuli in later phases of the experiment. In a preliminary study which employed the same stimuli as used here but in which there were provided only 40 response categories, such vocabulary exhaustion was found for a few *Ss*. Therefore, the larger number of response terms was used in the present experiment.

To investigate the effects of context on judgments, two experimental groups were employed, differing in initial context and subsequent shift. Table 1 indicates the design of stimulus presentation for the Low-High group. The reverse order was employed for the High-Low group. For purposes of analysis, the presentation of the notes was organized into "trials" or sets of five notes, within which each note of a current context was presented once. The arrangement of notes within the trials followed two 5 × 5 Latin squares to balance out the specific note-to-note sequences. The *Ss* were, of course, given no hint as to the grouping of stimuli into trials and phases. The transitional Phase 2 was

TABLE 1

SCHEDULE OF STIMULUS PRESENTATION FOR THE LOW-HIGH GROUP

	Trials	Numbers 29 Notes C♯ C.P.S. 135.6	32 E 161.3	35 G 191.8	38 B♭ 228.1	41 C♯ 271.3	44 E 322.6	47 G 383.7	50 B♭ 456.4	53 C♯ 542.8
	1	1	2	5	3	4				
	2	7	8	6	9	10				
	3	13	14	12	15	11				
	4	19	20	18	16	17				
Phase 1	5	25	21	24	22	23				
	6	29	28	30	27	26				
	7	35	34	31	33	32				
	8	36	40	37	39	38				
	9	42	41	43	45	44				
	10	48	47	49	46	50				
	11		51	52	55	53	54			
	12		57	58	56	59	60			
Phase 2	13			63	64	62	65	61		
	14			69	70	68	66	67		
	15				75	71	74	72	73	
	16				79	78	80	77	76	
	17					81	82	85	83	84
	18					87	88	86	89	90
	19					93	94	92	95	91
	20					99	100	98	96	97
	21					105	101	104	102	103
Phase 3	22					109	108	110	107	106
	23					115	114	111	113	112
	24					116	120	117	119	118
	25					122	121	123	125	124
	26					128	127	129	126	130

introduced to remove any abrupt shift which might be noticed by *S*s and generate hypotheses or suspicions about *E*'s purpose. To allay the feelings of inadequacy of the particularly naive and to help avoid the vocabulary exhaustion that might occur if the initial reference point was near one end of the scale, a middle *C* so designated was presented before the judgment series began. The experiment took 12.5 min. of actual judging time. In all, 36 Low-High *S*s and 37 High-Low *S*s provided complete protocols. Thirteen *S*s were eliminated because of incomplete records, failure to follow instructions, or getting out of step with the recording. None of the *S*s had to be eliminated because of vocabulary exhaustion. Because initial inspection of the data indicated that effects were greatest for those *S*s with least musical training, and because the High-Low groups contained fewer of untrained and more of highly trained persons than did the Low-High group, all analyses unless otherwise noted have been based upon two groups of 29 each matched for musical training.

In certain of the analyses, degree of musical training has been treated as a variable. For this purpose, a graded series of six categories has been employed. A score of one represented no musical training. A score of three represented the attainment of at least one of these: 3–5 yr. of choir, 1 yr. of private voice instruction, or 1–2 yr. of piano or instrumental instruction. A score of six represented the attainment of 8+ yr. of piano or other instrument, and/or study of theory, composition, orchestration etc. at the college level. Similar standards were used for the other categories, and each *S* was given the highest score for which he qualified.

RESULTS AND DISCUSSION

Context-induced shifts in response to the same stimulus.—The more usual analysis in terms of the point of subjective equality (PSE) has not been employed here. At best, it has a somewhat arbitrary meaning when used in the method of single stimuli except for those end-organ adaptations in which there is a qualitative

discontinuity in experience (e.g., the dimensions of hot-cold and sweet-salty). Where a situationally relative response language is employed, as high-low, heavy-light, or big-small, the PSE has at least a referent in the judgment language. In the present system of response categories, however, even this is lost. A similar analysis could be made by asking, for example, what note appears to *S* to be middle *C* at each stage of the experiment, but there would be no real reason to pick one reference note over any other. The typical PSE analysis looks at the shifting stimulus values corresponding to a constant response. The present study presents the data in a converse way, looking at the shifting response values corresponding to a constant stimulus. For the purpose of such analysis, this experiment was designed so that stimuli of value 41 (middle *C*♯) were present in each "trial" of all phases of the experiment, providing a common reference point against which to document the context effects. Figure 1 shows the course of judgments of this stimulus throughout the experiment for both the Low-High and High-Low groups. It is apparent that in general the traditional context effects resulting in a contrast illusion have been found. The following analyses spell this out in detail.

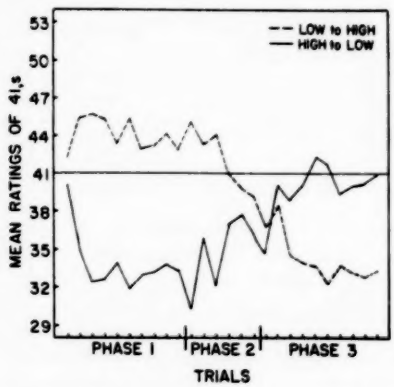

FIG. 1. Mean ratings of Note 41 throughout the 26 trials.

As a summary treatment of the effects of Fig. 1, two scores were computed for each *S*, one representing the mean of his judgments of Note 41 in Phase 1, the other the mean of 41's for Phase 3. These were then analysed as a 2×2 Latin square in a cross-over design (**4,** pp. 115–116). A highly significant effect of context was obtained. The *F* was 27.93 with 1 and 56 *df*, where an *F* of 7.12 would have been significant at the .01 level. Phases provided a nonsignificant *F* of 1.61, indicating that the general downward drift for an average of the two curves apparent in Fig. 1 does not reach significance. Groups (i.e., High-Low vs. Low-High marginals summed over both phases) provided a nonsignificant *F* of 2.39. Subjects provided an insignificant *F* of .78, indicating the absence of individual differences for mean judgments of both phases combined, reflecting, no doubt, the effect of the example middle *C* provided before the start of the judgments.

Inspection of individual trends seemed to indicate that those with less musical training were more susceptible to the context effect. To check on this, the Phase 1 data from the High-Low and the Low-High groups were given Treatments (high context vs. low context) by Levels (degree of musical training) analysis (**11,** pp. 121–151). The expected effect would appear as a significant interaction between Treatments and Levels. The *F* ratio for the interaction was but 2.62 where 3.18 would be required for significance at the

.05 level. (The relationship between degree of training and amount of Phase 1–Phase 3 shift is provided in a later section of the discussion.)

The clear-cut contrast effect obtained is in agreement with the general findings of the method of single stimuli for changes in stimulus context. Contrast effects thus cannot be interpreted as specific to the situationally relative, *E*-defined, and restricted-response language usually employed. The effect is in the direction predicted by Helson's adaptation level theory. However, for lifted weights, Helson (7, pp. 16–17) found that it was appropriate to disregard the temporal order of stimuli within the experimental period, and to treat earlier and recent stimuli equally in computing the adaptation level for any moment. While the present mode of analysis does not lend itself to the direct application of Helson's formula, it can be noted that an assumption of equal temporal weighting would predict that the curves in Fig. 1 would gradually converge in Phases 2 and 3, touching at Trial 26 when both groups had experienced the same total stimulation, and when Note 41 was being judged against the same adaptation level. The crossing of the curves indicates that recent stimuli are contributing much more heavily to the judgmental base than are more remote ones, even within the 12-min. experimental period. The t value for the difference between the two groups at Trial 26 is 2.74, $P < .01$.

Assimilation effects.—While the predominant trend is certainly toward the contrast illusion, there are nonetheless some *S*s whose records seem to indicate a systematic response shift in the other direction, i.e., to what has traditionally been called an assimilation illusion. For these *S*s Note 41 appears higher when it is in a context of *high* notes. Take, for example, the *S* who made the following judgments of Note 41 when it was the highest note in the low context in Phase 1: 40, 66, 35, 49, 46, 44, 44, 34, 45, 43; and these judgments when it was the lowest note in the high context of Phase 3: 76, 73, 64, 77, 78, 69, 64, 58, 66, 51. While such an individual record may appear convincing, there is a statistical problem in demonstrating that the assimilation effects are, over all, more than random deviations from a true value of no effect at all. If, in the population, true values ranged from extreme contrast to no illusion, "errorful" obtained scores would certainly fluctuate occasionally into the assimilation area. Such an interpretation would be in the interests of economy in psychological speculation, for to have in one setting two different illusion patterns in opposite directions would complicate judgment theory considerably. As the initial step in the evaluation of significant idiosyncrasy, separate analyses of variance were computed for each of the two experimental groups, in which *S* was used as one classification criterion and Phase 1 vs. Phase 3 as the other, the 10 judgments of Note 41 being the cell entries. If some *S*s reacted systematically differently to the shift in context than did others, this would appear as a significant *S* by Phase interaction. For the Low-High group, this interaction provided an F ratio of 18.0 where 1.80 would be significant at the .01 level. For the High-Low group, the F was 16.2. Such a significant interaction does not, of course, prove that there were some significant assimilators—a significant interaction could have occurred had some *S*s shown very slight contrast effects, others very large ones, all in the same direction. But it does legitimize inspecting individual cases. For this purpose, a t ratio was computed for each assimilating *S* between his 10 judgments

of Note 41 in Phase 1 and his 10 judgments in Phase 3. As one would expect, many of these t values are low. However, when one considers the relatively small number of observations that were involved, some of these values are strikingly high, the highest being 5.62. Of the 36 Low-High *Ss*, t values for differences in the assimilation direction in excess of the .01 level value of 2.552 were obtained for 3 *Ss*. The likelihood of such an event occurring by chance is but .006. Of the 37 High-Low *Ss* there were also 3 assimilation t's in excess of the .01 level, for which $P < .007$. These t's which are beyond the .01 level of 2.552 seem clearly more frequent than could be accounted for by chance variations around a true value of no difference. When tested by the frequency of instances in excess of the .05 level ($t = 1.734$), however, the departure from chance is not significant. For the Low-High group there were four such cases ($P < .11$) and for the High-Low group there were four cases ($P < .12$). While the evidence from the .01-level analysis seems compelling, all in all we should probably hold in abeyance judgments as to the reality of assimilation errors until more data are at hand.

The willingness of the present writers to accept the existence of assimilation errors in these data is increased by the discovery of such effects in more substantial number in another judgmental setting (**3**), although quite different stimulus material and response scale were used. If there were assimilators, why? One guess is that such persons had good enough pitch constancy to notice the shift in general stimulation level but over-reacted to this awareness. It is regretted that introspective accounts were not collected from the *Ss* at the termination of the experiment. It is possibly relevant that the extreme assimilators came from the middle levels of musical training rather than from among the untrained.

Effect of context shift on fineness of discrimination.—In the method of single stimuli, with the usual limited number of response categories, a marked shift of the stimulus context or the introduction of a remote anchor has the effect of reducing the fineness of differentiation (**14**). The usual analysis of such effects in terms of coarsening of judgment categories can also be studied as a reduction in the standard deviation of judgments made, regardless of accuracy. Another measure of fineness of discrimination is provided by the correlation of the judgments with the true stimulus values. Both measures were employed, and, as entirely similar results were obtained, only the analysis of correlations will be described in detail.

For each *S* a correlation of all 50 judgments with true stimulus values was computed separately for Phase 1 and Phase 3. Transformed into z' values, these correlations have been used as scores and analyzed in a cross-over design (**4,** pp. 115–116). The logic of the design and the mean value for each cell and marginal are shown in Table 2. For phases the F is nonsignificant ($F = 3.02$, $P > .05$).

TABLE 2

EFFECT OF PHASE AND CONTEXT ON CORRELATION OF JUDGMENTS AND STIMULUS VALUES (MEAN r's)

Context	Phase 1	Phase 3	Mean
Low	.79 LH	.79 HL	.79
High	.85 HL	.84 LH	.85
Mean	.82	.81	

Group Means: LH (Low-High) = .81; HL (High-Low) = .82

The effect of context is highly significant ($F = 26.98$, $P < .01$). The group difference produces an $F < 1.00$. Individual differences are significant, $F = 5.07$, $P < .01$. The failure to find significant effects for phases indicates that the shift in context has resulted in no loss of discrimination. This result differs from the typical finding for shifts in stimulus context, as through the introduction of extreme anchors, or as in the previously cited study by the present authors (**3**) in which the same general experimental design and analysis were used but with a typically restricted range of response categories. Tentatively, the difference is attributed to the relatively unlimited range of judgment categories available, obviating the necessity for the *S* to readjust category usage when the shift in stimulation level occurred.

The highly significant effects of context indicate the departure of the note scale from Weber's law, typically found for this range of frequencies (**6**). Because of the symmetrical nature of the design, it is believed that this factor is not confounded with any of the other experimental effects studied, although in a detailed application of an adaptation-level prediction to such data, this complication would have to be included.

Correlations between prior training and other variables.—The amount of prior training has considerable relevance from a standpoint of adaptation-level theory, although amount of training is no doubt confounded with whatever innate capabilities exist in this area, being in part a *symptom* of talent as well as a *cause*. For this purpose the degree of training category numbers 1 to 6 were used as measurement values, and product-moment correlation coefficients computed separately for the Low-High and High-Low groups. The total *N*'s of 36 and 37 were employed.

As would be expected from a presumably more stable pre-experimental adaptation-level, the more highly trained *S*s show less Phase 1 to Phase 3 shift in their judgments of stimuli of value 41 (r's = $-.32$ and $-.34$), are more accurate in their judgments of Phase 1 stimuli as measured by z' (r's = .20 and .53), and have smaller *SD*'s for their Phase 1 judgments (r's = $-.51$ and $-.49$). It might have been expected, if accuracy of judgment were a function of a stable adaptation-level, that those with least training would profit most from the experience within the experimental session and would show the greatest gain in accuracy (measured as Phase 3 z' minus Phase 1 z'). This, however, was not found (r's = .09 and .05). Other correlations computed show the Phase 1 *SD*'s highly related to the magnitude of the Phase 1–Phase 3 shift in judgments (r's = .80 and .58), and relatively independent of the z' measure for Phase 1 (r's = $-.07$ and $-.29$). Accuracy of discrimination in Phase 1 as measured by z' shows little relationship to extent of judgment shift (r's = .27 and $-.01$).

A word might be said about *S*s' claims of absolute and relative pitch. The one student claiming absolute pitch provided an excellent, although not perfect, performance (e.g., mean of judgments of Note 41 : Phase 1 = 40.0 and Phase 3 = 40.0; r's = .90 and .99; *SD*'s 4.38 and 4.00 where the true *SD* was 4.24). No *S*s showing perfect pitch were found. The claim of relative pitch made by 27 *S*s indicated nothing in the way of superior performance on this task.

Summary

In the typical studies using the method of single stimuli, the judgment language provided is situationally relative, an arbitrary creation

of the *E*, and restricted in number of response categories. These specific features could account for the typical context effects found, without a distortion in identity judgments being involved. To check on this possibility, the present study introduced a response language anchored outside of the experiment, absolute in that it referred to invariant aspects of the stimulus, and extensive in the number of categories available.

The *S*s were asked to judge the pitch of notes in terms of a piano keyboard reference system. The Low-High group first heard notes varying in a relatively low context and later were shifted gradually to a high context. The High-Low group received the reverse treatment. The traditional adaptation-level or contrast illusion was found in the shifting judgments of a common note repeated in both contexts, showing this effect to be independent of the specific semantic details of the usual method of single stimuli assessment. However, the usually found loss of refinement of discrimination consequent to a shift in context was not obtained, and presumably such effects are a result of the restricted number of response categories typically provided. In addition to the typical contrast effects, a small, but possibly significant, number of *S*s seemed to respond to the shift in context with an assimilation error. Persons with advanced musical training were less influenced by a shift in context and showed more accuracy in discriminating among the stimuli presented.

References

1. Bevan, W., & Darby, C. L. Patterns of experience and the constancy of an indifference point for perceived weight. *Amer. J. Psychol.*, 1955, **68**, 575–584.
2. Bressler, J. Judgment in absolute units as a psychophysical method. *Arch. Psychol.*, 1933, **23**, No. 152.
3. Campbell, D. T., Hunt, W. A., & Lewis, N. A. The effects of assimilation and contrast in judgments of clinical materials. *Amer. J. Psychol.*, 1957, **70**, 347–360.
4. Cochran, W. G., & Cox, G. M. *Experimental designs.* New York: Wiley, 1950.
5. Eriksen, C. W., & Hake, H. W. Absolute judgments as a function of stimulus range and number of stimulus and response categories. *J. exp. Psychol.*, 1955, **49**, 323–332.
6. Harris, J. D. Pitch discrimination. *J. acoust. Soc. Amer.*, 1952, **24**, 750–755.
7. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
8. Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
9. Hunt, W. A., & Volkmann, J. The anchoring of an affective scale. *Amer. J. Psychol.*, 1937, **49**, 88–92.
10. Johnson, D. M. Learning function for a change in the scale of judgment. *J. exp. Psychol.*, 1949, **39**, 851–860.
11. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
12. Parducci, A. Direction of shift in the judgment of single stimuli. *J. exp. Psychol.*, 1956, **51**, 169–178.
13. Woodworth, R. S., & Thorndike, E. L. Judgments of magnitude by comparison with a mental standard. *Psychol. Rev.*, 1900, **7**, 344–355.
14. Volkmann, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951, 273–294.

(Received February 4, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

TIME SHARING BETWEEN FREQUENCY-CODED AUDITORY AND VISUAL CHANNELS [1]

E. T. KLEMMER [2]

Operational Applications Laboratory, Air Force Cambridge Research Center

An earlier study investigated the effect upon performance of rapid shifts of attention between a visual information source and an auditory information source (**2**). The visual channel utilized position-coding for both stimuli and responses and was considerably easier than the auditory channel which used frequency-coded stimuli and position-coded responses. From the results of this earlier study it was hypothesized that relative degradation due to time-sharing is correlated with differences in channel difficulty. Specifically, it appeared that the easier channel suffers more from forced alternation than the more difficult channel.

The present study was designed to provide additional evidence on this hypothesis as well as to provide information about other pertinent questions concerning performance in a shared-time situation. The auditory and visual sources were selected to be about equally difficult when presented separately. The rate of alternation between channels was varied as before. In addition, a test forcing random alternation between channels was compared to regular alternation at the same mean rate. Finally, the effect of presenting redundant stimuli simultaneously on both channels was investigated.

[1] This research was performed at the Operational Applications Laboratory, Bolling Air Force Base, Washington, D. C., in support of Project 7682 and is referenced AFCRC TR 57-1, ASTIA Document No. AD 110061.

[2] Now at IBM Research Center.

METHOD

Apparatus.—The *S* sat in a small, dimly lighted room isolated from the programming apparatus. He held his thumb, middle finger, and small finger on three keys during each run. Immediately above the keys was a 1-in. circular milk-glass screen upon which the colored-light stimuli were projected with even illumination so that no position cues were present. Each visual stimulus was chosen from a class of three colors: red, green, and orange. The *S* pressed keys in one-to-one correspondence with the color of the light stimulus: red—left key; orange—center key; green—right key. The three colors were approximately equated for brightness by successive comparison judgments. Spectral distributions of the colors were not determined.

The auditory stimulus consisted of a sine-wave tone of either 100, 700, or 5000 cps presented over a loudspeaker. The three tones were approximately equated for loudness by successive comparison judgments. The *S* responded to the auditory stimulus by pressing one of the same set of keys used for the visual channel. The low-frequency tone corresponded to the left key, middle tone to the center key, and high tone to the right key.

The tones and lights were programmed by teletype tape at a constant rate of 2 per sec. with a 50% on-off ratio in all tests. The *S*'s responses were held in a relay memory circuit and punched onto teletype tape at the end of each cycle. The *S* did not have to respond to one stimulus before the occurrence of the next stimulus in order to be scored correct. Rather, a one-cycle scoring interval was centered about *S*'s own mean reaction time from each stimulus. Any response within this one-cycle interval was scored as a response to that stimulus. For example, the scoring interval might be set to include all responses from .25 sec. to .75 sec. from stimulus onset. Response times were accumulated on a Standard Electric timer from which reaction times were computed after each run. The scoring interval was located according to each *S*'s own previous reaction time on that test.

Subjects and training.—Eight university students served as paid *S*s, but not all *S*s served

in all test series because the full study ran over many weeks and some *Ss* left the program during that time. Each *S* was given two to five weeks of training, one hour per day, five days per week. The differences in training were due to *S* availability during the pilot runs. The training included the auditory and visual channels separately and in alternation. This training was not sufficient to provide unchanging performance so each test series included its own control tests.

Regular alternation tests.—The regular alternation series consisted of four different tests: two with alternation between the auditory and visual channels plus one control test on each channel separately. One alternation test changed channels after each stimulus (Test 1) and the other presented five stimuli in a row on each channel before changing to the other channel (Test 5). Thus, Test 1 alternated between channels every .5 sec. and Test 5 alternated every 2.5 sec.

Each run with alternating channels contained 150 stimuli on each channel with equal frequencies of occurrence of all stimuli, i.e., each color and each tone occurred 50 times. The order of stimuli within each channel was random. Each run on the visual channel separately (Test V) or the auditory channel separately (Test A) consisted of 150 stimuli with each color or tone occurring an equal number of times in random order. Thus, the runs on the alternating channel tests lasted 150 sec. and the runs on the single channel tests were 75 sec. long.

During the alternation tests, there were no overlapping stimuli or breaks in cycling when the channel was changed. Thus, a stimulus occurred every .5 sec. in one channel or the other, but not both at once.

Seven *Ss* took single runs on each test in the following order: V, A, 5, 1—and repeated this order 20 times so that each *S* took 20 runs on each test. Two *Ss* ran alternately each session so that about 10 runs were completed in an hour. Five different program tapes were used in each test to prevent learning of the particular random sequences employed.

Random alternation test.—As the name implies, the random alternation test (Test R) employed unpredictable alternation between the auditory and visual channels. In this test, the number of consecutive stimuli on one channel could be any number from 1 to 10, inclusive, before changing to the other channel. The number of consecutive stimuli was simply selected from a table of random numbers. The order of stimuli on each channel was random with only the restriction that the total number of responses on each key be equal in each run. As in all tests, the same keyboard was used for both channels.

Each run of the random alternation test presented a total of 150 stimuli at the cycling rate of 2 per sec. The number of stimuli on each channel varied somewhat from 75 from run to run because of the randomness of alternation. Seven *Ss* took five practice runs followed by 15 test runs. Between each set of five random alternation runs, each *S* took one run of Test 5 of the regular alternation series as a control test. This control test will be designated Test 5r. Test 5r changes channels regularly every five stimuli which is approximately the mean rate of alternation in the random test.

Redundant-simultaneous test.—Many studies have found that simultaneous visual and auditory presentation of redundant material leads to better performance (1). The redundant-simultaneous test (Test R-S) was designed to determine the improvement that results from presenting the colored lights and tones simultaneously when both stimulus channels call for the same response.

The cycling rate was again 2 per sec. and each run contained 150 randomly ordered stimuli. During each .5 sec. of the run there was both a light and a tone stimulus which started and ended together. The stimulus-response correspondences were the same as before: red—100 cps—left key; orange—700 cps—center key; green—5000 cps—right key.

Five *Ss* took one practice run and five test runs on the redundant test. Between runs each *S* took one run on each channel separately as control tests. These control tests will be designated Test Ars and Test Vrs.

Results and Discussion

The response tapes were scored for the number of correct responses in each run of each test. For each *S*, the number correct was averaged over repeated runs of the same test and the mean number correct converted to a percentage. In all tests there was improvement with practice over the repeated runs. The average percentage correct score went up about 10% between the first and last runs of all tests and all *Ss*. It was for this reason that control condition test runs were intermixed with runs on the experimental tests. Determination of statistical signifi-

TABLE 1

PERCENTAGE CORRECT SCORES IN KEY PRESSING TO FREQUENCY-CODED LIGHTS AND TONES AT TWO PER SECOND

S	Regular Alternation Series						Random Alternation Series				Redundant Series		
	Alternation Tests				Control		Random		Control		Redundant	Control	
	5		1		V	A	R		5r		R-S	Vrs	Ars
	Lights	Tones	Lights	Tones	Lights	Tones	Lights	Tones	Lights	Tones	Both	Lights	Tones
B	80	73	46	37	83	91	86	86	89	84	96	81	94
D	93	85	86	76	96	91	94	96	95	96	96	97	98
G	81	69	43	30	87	85	83	83	86	77	95	86	87
K	79	69	48	47	84	81	88	88	92	81	93	88	88
L	98	95	53	55	98	97	96	88	98	95	—	—	—
M	80	76	53	47	77	93	85	86	95	87	96	69	97
R	—	—	—	—	—	—	84	74	80	68	—	—	—
W	97	96	89	87	98	96	—	—	—	—	—	—	—
Mean	87	80	60	54	89	91	88	86	91	84	95	84	93

cance within data of a single *S* was based on the distribution of differences in scores from paired runs—one control and one experimental run. The *SD* of the differences between the paired percentage correct scores used below ranged from 2% to 11% over *S*s. Averaged variance leads to an *SD* of 5% for these same comparisons. The same method was used for significance tests for the reaction time data, even though reaction time changed little with practice. The *SD* of the differences between paired reaction times used below ranged from .017 sec. to .079 sec. over *S*s. Averaged variance leads to an *SD* of .033 sec. for the same comparisons.

Table 1 presents the percentage correct scores for each *S* and each test. The implications of these scores are discussed below under separate headings. Mean reaction times were computed for each run of each test from the clock readings. These reaction times, averaged over runs, appear in Table 2. The questions which can be answered from the data of the present study will now be considered.

Are frequency-coded lights or tones more difficult?—The answer to this question is in Table 1. The lights presented separately averaged 89% correct; the tones, 91% correct. This small advantage of the tones in the average data for seven *S*s is not supported by a look at the individual *S*s. Five of the seven *S*s did better on the lights and only two on the tones. The average difference between channels is thus of no consequence, but is the difference within single *S*s important? To answer this question a *t* test was applied to the distribution of differences in number-correct score between paired runs, one on each channel. Of the five *S*s who performed better on the light channel, two (L and D) did so with a difference significant at the .05 level. The two *S*s who did better on the tone channel (B and M) did so with a statistical significance beyond the .01 level. Thus, performance on the separate auditory and visual channels is really different for some *S*s; but five of the seven *S*s had scores within 5% of each other on two channels.

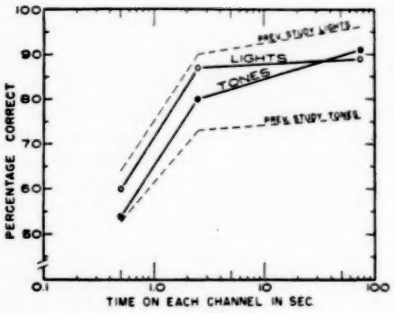

FIG. 1. Mean percentages of correct responses to lights and tones for seven *S*s as a function of time on each channel in seconds.

How does rapid alternation between channels affect performance?—The earlier study had indicated little decrement in performance until alternation occurred faster than once every 2 sec. and a rapid drop-off thereafter (**2**). The pertinent data from the regular alternation series of the present study are plotted in Fig. 1. The average percentage correct score for seven *S*s is plotted against time on each channel. Note that the time scale is logarithmic with the separate channel tests plotted at 75 sec., which was the length of each run in Tests A and V. For purposes of comparison, the results from the three *S*s of the previous study are drawn as broken lines.

The finding that there is a small decrement in performance due to alternation at rates less than once every 2 sec. is sustained. The large drop at higher rates of alternation is also found again. In analyzing Fig. 1, two differences between the studies should be remembered. The previous study used position coding on the light channel and gave all runs of one test together starting with the separate channel tests. The tone channel, however, was identical in the two studies so that the large difference in scores must be attributed to individual differences in *S*s and training.

Which channel suffers more from alternation?—The previous study had suggested that the easier channel was hurt more by alternation. This finding is evident in the dashed lines of Fig. 1 but was much more striking in tests conducted at 3 per sec., which are not shown. The present tests employ separate channels of approximately equal difficulty; therefore, if one applies the rule that the easier channel suffers more, it would be predicted that both channels would show about the same decrement due to alternation. Figure 1 bears on this point. It shows a larger drop in the tone channel than in the light channel due to alternation. Indeed, the light channel dropped only 2% between Test V and Test 5 while the tone channel dropped 11%. The difference between 2% and 11% is significant at the .02 level. The 2% drop in the light channel does not reach the .05 level of significance. Six of the seven *S*s dropped more in the tone channel than in the light channel in going from separate presentation to alternating every 2.5 sec. The seventh *S* (W) showed virtually no change in either channel. Thus, the tone channel is generally hurt more by alternation even for those *S*s who did better on the light channel when presented singly. These results suggest that the auditory channel is more vulnerable than an equally difficult or slightly easier visual channel. The earlier study showed that the auditory channel was less vulnerable than a much easier visual channel. So we now have two hypotheses: that the easier channel is hurt more by alternation and that the auditory channel is hurt more by alternation. Mowbray (**3**) obtained similar results in a very different situation. He used simultaneous visual and auditory presentation of different prose passages and found that the auditory channel was disrupted more than the visual and that the easy material was disrupted more than the difficult when compared to separate presentation.

Returning to the results of the regular alternation tests, one can see that the additional drop from alternation every 2.5 sec. to alternation every .5 sec. is just about the same on each channel

TABLE 2

REACTION TIMES IN KEY PRESSING TO FREQUENCY-CODED LIGHTS AND TONES AT TWO PER SECOND

(All table entries are in units of .01 sec.)

S	Regular Alternation Series				Random Alternation Series		Redundant Series		
	Alternation Tests		Control		Random	Control	Redundant	Control	
	S	I	V	A	R	Sr	R-S	Vrs	Ars
B	57	58	60	52	54	54	45	57	49
D	55	57	54	56	54	53	53	52	55
G	51	46	46	49	51	54	47	45	52
K	47	50	45	48	48	47	42	45	49
L	54	52	53	51	51	51	—	—	—
M	50	47	45	43	53	55	43	43	42
R	—	—	—	—	48	49	—	—	—
W	51	51	50	48	—	—	—	—	—
Mean	52	52	50	50	51	52	46	48	49

when performance is measured in percentage correct. Table 1 shows that all *S*s show about equal drop in both channels; this drop averaging 26%.

How does performance under random alternation compare with regular alternation at the same mean rate?—The random alternation series was designed specifically to see how much more difficult would be *S*'s task when he could not predict the time of channel shifts. The findings appear in Table 1. The average for seven *S*s showed the tone channel 2% better during random alternation and the light channel 3% better during regular alternation. Looking at scores of individual *S*s, one can see that these small differences are not consistent over *S*s, so the conclusion must be that random alternation produced the same level of performance as regular alternation at the same mean rate.

Can a gain in performance be expected as a result of redundant-simultaneous stimulus presentation?—There is evidence of an improvement in performance due to combined visual and auditory presentation in Table 1. That is, for the five *S*s who took Test R-S, the percentage correct score for the redundant-simultaneous presentation was 95% compared to 84% and 93% for the control tests using separate presentations. Only the difference between 95% and 84% was significant and this at the .05 level. If the scores for individual *S*s are studied there is a suggestion of something more than a general improvement due to simultaneous presentation. In Table 1 the one *S* (D) who performed nearly perfectly on the separate channels actually showed a little lower performance on the combined presentation. The two *S*s (G and K) who performed below 90% and about equally well on the separate channels showed 6% improvement in going to the combined presentation. The two *S*s (B and M) for whom the separate channels were most different, performed little or no better on the combined presentation than the better single channel. Statistically speaking, the redundant-simultaneous presentation was significantly better than the better single channel at the .05 level only for the *S*s G and K. No other differences approached significance. Several hypotheses might explain these results, one of the more interesting being that the redundant multichannel presentation is effective only if the separate channels are nearly of the same difficulty.

How do reaction times compare for frequency-coded light and tone stimuli?—Table 2 shows that the average reaction times are equal for the light and tone channels when presented separately. Thus, the channels, which were about equal in difficulty when averaged over *S*s, also show equal average reaction time. It may also be asked whether reaction time is correlated with performance within the individual *S*s. The answer to this question is that, of the four *S*s who showed more than 2% difference in performance between the two channels, all had a lower RT on the better channel. A *t* test based on the distribution of differences between paired runs shows significantly different mean reaction times for all four of these *S*s (.01 level). However, the difference in reaction times between channels is only .03 sec. or less for six out of the seven *S*s, so that the general finding is still that frequency-coded lights and frequency-coded tones showed about the same reaction time as well as the same level of difficulty.

How does alternation between channels affect reaction time?—The effect of alternation between channels upon reaction time is small but consistent. From Table 2, it can be seen that the average reaction time for alternation tests 5 and 1 is .52 sec. (Separate reaction times for each channel could not be measured in the alternation tests.) Both separate channel tests show a reaction time of .50 sec. The difference between mean RT on Test 5 and mean RT of Tests V and A is significant at the .02 level. The average reaction time did not change in going from alternating every 2.5 sec. (Test 5) to alternating every .5 sec. (Test 1) and the individual *S*s showed inconsistent differences. The answer to the question, then, is that reaction time increases slightly with alternation and is not very sensitive to rate of alternation.

The previous study employed a visual channel with a lower reaction time than the auditory channel and found that, during alternation, reaction time was about equal to larger reaction time of the single channel tests (**2**).

Does the redundant-simultaneous presentation produce a shorter reaction time than the separate channel tests?—The reaction times for Test R-S in Table 2 show that the redundant presentation had a mean RT .01 sec. less than either separate channel control test, but three of the five *S*s had a lower reaction time on one of the single channels, so that it must be concluded that there is no evidence that the redundant-simultaneous presentation leads to a reaction time different from the faster separate channel.

Summary

Performance on a high-speed key-pressing task was determined for situations in which *S*: (*a*) responded to a frequency-coded visual channel (color-coded); (*b*) responded to a frequency-coded auditory channel (tone-coded); (*c*) alternated regularly between the visual and auditory channels; (*d*) alternated randomly between the visual and auditory channels; and (*e*) responded to simultaneous visual and auditory channels presenting redundant stimuli. Mean reaction times were taken during all tests.

Each of the above tests was given to 5 to 7 trained *S*s with the following results:

1. The frequency-coded lights and tones were of approximately equal difficulty; two *S*s doing significantly better on the tones, and two on the lights, and three showing insignificant differences.

2. Alternating between channels every 2.5 sec. resulted in little decrement in the light channel. The tone channel dropped 11% in percentage correct score, the two *S*s who did better on the separate tone channel showing the largest drop.

3. Increasing the alternation rate to once every .5 sec. dropped percentage correct performance on both channels an additional 26%.

4. Random alternation between channels produced the same performance as regular alternation at the same mean rate.

5. Presenting redundant information in the two channels simultaneously resulted in higher performance than the better single channel for only two of five *S*s.

6. The average reaction time was the same for the separate light and tone tests. Alternation between channels increased the reaction time but only by .02 sec. The redundant-simultaneous presentation gave a reaction time no different from the single channel which had the shorter reaction time in the separate tests.

References

1. Day, W. F., & Beach, B. R. A survey of the research literature comparing the visual and auditory presentation of information. Air Materiel Command, United States Air Force Tech. Rep. No. 5921, Nov. 1950.
2. Klemmer, E. T. Time sharing between auditory and visual channels. In G. Finch and F. Cameron (Eds.), *Air Force human engineering, personnel, and training research*. Washington: National Academy of Sciences—National Research Council, Publ. No. 455, 1956. Pp. 199–203.
3. Mowbray, G. H. Simultaneous vision and audition: the comprehension of prose passages with varying levels of difficulty. *J. exp. Psychol.*, 1953, **46**, 365–372.

(Received February 15, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

A CORRELATIONAL STUDY OF TWO REASONING PROBLEMS

LARRY BRUNK,[1] E. GORDON COLLISTER,
CAROLYN SWIFT, AND SAM STAYTON
University of Kansas

For half a century the difficult reasoning problem individually administered to the human *S* has been studied in the psychological laboratory, and major theoretical interpretations of thinking have relied heavily on such studies (e.g., **5, 6, 13, 14**). The purpose of this study is to explore a heretofore neglected aspect of problem solving—the consistency of individual performance on a series of difficult problems. Does *S*'s performance on two or more problems correlate significantly? What are some limiting conditions of whatever consistency is found? Previous studies largely ignore these questions.

The sort of experimental task here considered is difficult; its solution demands a new integration of familiar concepts. Neither memorization nor the acquisition of a skill is involved. Instead, there is a principle to discover, and once this principle is found the problem is solved. Following Maier (**9, 10**), such a task is designated a reasoning problem. Ray (**11**) has given several examples of what are here called reasoning problems.

The literature on problem solving shows little concern with the consistency of individual performance. Most studies involving the administration of two or more problems to the same *S*s do not give correlation coefficients or data from which they can be computed (e.g., **4, 6, 12**). However, two recent studies do give such data and the relationships reported are low. Adamson (**1**) administered three of Duncker's problems to the same *S*s and found only one out of three proficiency relationships significant at the .05 level. Guetzkow (**7**) correlated proficiency on Maier's pendulum problem with a composite score derived from several similar tests. Using more than 200 *S*s, a low positive, but significant, correlation was found (r point biserial = .34, our calculations). From these data it appears reasonable to infer a low correlation between performance on successive reasoning problems. The theory presented below accounts for this low correlation, and, in addition, suggests conditions under which correlations of proficiency may be high.

A theory of reasoning proficiency.—It is postulated that at least two abilities are involved in problem-solving behavior. One is the ability to conceptualize the solution to a given problem. This ability is assumed to be stable provided there is sufficient similarity in the problems to be solved. The other is the ability to reject erroneous hypotheses, which is held to be highly variable. The two abilities are held to be essentially uncorrelated.

The variability of the hypothesis-rejecting process stems in part from the structure of the reasoning problem. By definition, several apparent paths to the goal are available to *S*, and it is a matter of chance (from *E*'s point of view) whether or not he selects the correct one. If he should choose an incorrect path, an inhibitory process may occur:

[1] Now with U. S. Naval Air Development Center, Johnsville, Penna.

preoccupation with an incorrect lead results in inhibitory perseveration, which temporarily lowers *S*'s ability to shift to the correct solution and thus solve the problem. These two factors, the probability that chance will determine whether the initial selection is the correct one, and the inhibitory function of perseveration due to an erroneous hypothesis, are held to account for the high variability in performance on reasoning problems as customarily administered.

If this analysis is correct, an experimental technique which promptly identifies an hypothesis as being incorrect should produce a high intercorrelation between proficiency scores of two problems so administered. In this case the ability to conceptualize the solution would be measured, and this ability is assumed to be stable, given sufficient similarity in the problems used.

The same two tasks are used in each of the four experimental conditions to be described below. Condition 1 involves administering the tasks under optimal conditions according to the theory; the performance correlations should accordingly be high. In Cond. 2 one task is administered in a fashion calculated to arouse perseveration; this condition should yield a low correlation, and comparing Cond. 1 and 2 should test the theory outlined above. Conditions 3 and 4 explore the effects of two techniques commonly used in problem-solving experiments: in one, *S* is asked to "think aloud" as he solves one of the problems, and in the other, *S* is placed under mild time pressure.

General Method

Materials.—One of the two tasks was devised by Forrest Ladd, the other by Brunk (**2**). Both tasks are based on the Vigotsky test. Each contains 22 blocks which divide into four logical groups. In the Ladd Test, 11 blocks are whole figures (squares, circles, and octagons) and the remaining blocks are half figures (rectangles, half-circles, and half-octagons). Both of these groups are subdivided into blue and red colors. The four logical groups are: red wholes, blue wholes, red halves, and blue halves. An insightful solution by *S* takes both the color and shape divisions into account. No particular wording is necessary for the shape distinction to be considered insightful; if *S* can indicate that half the blocks are shaped somewhat like footballs and the remainder more like basketballs, the description is considered correct. Identifying names appear on the bottom of each block.

In the second task, the PAMS Test, two concentric shapes are painted on the top of each block (and on the bottom, together with an identifying numeral). The blocks are grouped according to the interrelationship of block shape and the two figures painted on the block. In Group I the two painted shapes are identical with the shape of the block (e.g., two circles are painted on a circular block). In Group II the outer painted figure is congruent with the block shape, while the innermost figure is different (e.g., the larger figure painted on a circular block is a circle, but the innermost figure is a triangle). In Group III the painted shapes are congruent but are different from the shape of the block (e.g., two concentric circles are painted on a hexagon). In Group IV all three shapes are different (e.g., a circle and a triangle are painted on a hexagon). A solution is considered insightful at the point when three of the four groups are accurately described.

Procedure

Condition 1[2]—The technique of administration of the tests considered standard in these experiments is the "one-block administration" here described. In this administration it was desired to give frequent and cumulative clues "optimally" distributed. This was done by making the clues an integral part of the tasks, and by allowing *S* to determine, by his own activity, the time interval between clues.

At the outset of testing, one of each of the four kinds of blocks is placed before *S* in each corner of the administration board which is a standard desk blotter. The four blocks are placed name up and represented as examples of the four kinds of blocks *S* will encounter. The *S* then obtains the remaining blocks one at a time from a cardboard slot. Without seeing the name, *S* must compare the block to those exposed on the board and place it in the corner in which he thinks it belongs. He then gets the clue by turning the block name up. If the

[2] We are indebted to Omer J. Rupiper for assistance with this part of the experiment.

name corresponds to that on the blocks thus far accumulated in that corner, he has placed it correctly. If the name is different, he replaces the block in the correct corner. He then obtains another block, continuing in this manner until all blocks are placed. The blocks remain on the board after each placement as cumulative clues to the principles involved.

The *S* is encouraged to take as much time as he likes both in placing the blocks and in verbalizing the principles. He is also assured that the blocks are divided by a logical plan, and that once this division is understood the remaining blocks can be correctly placed.

After placing Blocks 3, 8, 15, and 18 (last) *S* is urged to verbalize the principle. When all 22 blocks are exposed they are left before *S* until he produces what he considers to be a relevant generalization or indicates he is unable to do so. Finally, if he does not verbalize a solution he is specifically questioned (e.g., "How is this group different from that group?"). When *S* has produced the correct generalization or has exhausted his possibilities, the board is cleared and the next task administered. Both tests are given in the same sitting with the blocks of each task administered in an invariant sequence. Both test sequences are used.

Performance is scored on a 5-point scale, depending on the point at which *S* correctly verbalizes the principles. Correct verbalization on Blocks 1 to 7 is scored 1; on Blocks 8 to 13, 2; on Blocks 14 to 17, 3; Block 18, 4; and a score of 5 is given if no adequate solution is proffered. Thus a low score indicates a better performance.

Condition 2.—In the above-described group two similar tests were administered to all *Ss* with frequent and successive clues which should operate to identify incorrect hypotheses immediately. Under these conditions a high performance correlation would support the theory presented above. In the present group the same two tests are administered, one with the frequent clues, and the other with clues which are considerably more widely spaced, i.e., they occur at intervals which are greater than is believed optimum. Under these conditions, performance should correlate negligibly according to the theory. As noted, the two tests are quite similar; but if similarity is the crucial factor then both conditions should be characterized by the same performance correlation. If, on the contrary, Cond. 1 and 2 groups show significantly different performance correlations, the theory presented above will be supported.

The *Ss* in Cond. 2 were given the PAMS Test in the one-block administration and the Ladd Test in a Vigotsky-like administration to be described. Both test sequences were used in both groups.

In the Vigotsky-like administration it was desired to administer the test in a manner that would allow inhibitory perseveration. The construction of the tasks makes it impracticable to dispense with the clues altogether; accordingly it was decided to give clues at "greater than optimal" intervals. This was effected by requiring that *S* accomplish a certain amount of activity between clues. The procedure was adapted from the Hanfmann-Kasanin administration of the Vigotsky test. The *S* is presented all blocks simultaneously; he must group them into four categories and then explain his grouping before he is given a clue. Since this means an interval of a few minutes between each clue, the clue does not have the direct reinforcement it does in the one-block administration, in which a single choice is promptly indicated as right or wrong.

The Vigotsky-like administration of the Ladd Test was scored according to the total number of clues necessary prior to solution. As in the one-block administration, non-insightful solvers were given a tied score in last place. Since *Ss* required between 1 and 9 clues the score of 10 was assigned non-insightful solutions and the total range is from 1 to 10.

Condition 3.—This condition is concerned with investigating the effects of the injunction to "think aloud" on proficiency in solving reasoning problems. Such instructions have been used extensively in major theorizing (e.g., **5, 6, 14**). In this experiment the question is raised concerning the effect of the instruction to "think aloud" on performance: Does this request significantly affect proficiency?

The *Ss* were first given the Ladd Test in the standard one-block manner, and were then given the PAMS Test with the request they describe their thoughts as they worked. Each *S* was requested to tell why he placed each block where he did and was encouraged to verbalize freely, with *E* recording what he said. In other respects the administration and scoring were like the standard one-block procedure described above (Cond. 1). Verbalization is required in both procedures, but there is a considerable difference in the behavior demanded. In the regular administration *S* is asked at infrequent intervals to describe an objective stimulus configuration; in this condition *S* is urged to verbalize his thoughts continually as he works.

Condition 4.—In the previous conditions *Ss* were told that the time consumed was unimportant; to emphasize this they were informed that no time record was kept. Condition 4 explores the effects of placing the two tests

into the more traditional framework of a timed test.

The *S*s were treated as those in Cond. 1 but with two modifications. A stopwatch was on the table before *S*, and the initial instructions with regard to time were altered as follows: "This is timed, but time isn't important. The main thing is to tell me the principle." The second modification was that *S* was not reminded that time was unimportant if he appeared to be hurrying.

Subjects

The *S*s total 147, with 86 men and 61 women. About the same distribution by sex occurs in each of the subgroups used. Two sources of *S*s were used, students in introductory psychology courses and students reporting to the University of Kansas Guidance Bureau for counseling. The two sources yielded *S*s with comparable scores on the American Council on Education Psychological Examination, and on the two problems here used. Accordingly, the sources will not be differentiated in the following account.

Condition 1 had 67 *S*s; Cond. 2 had 40, and Cond. 3 and 4 had 20 each. The order of giving the two tests was rotated in Cond. 1, 2, and 4. In Cond. 3 *S* was asked to "think aloud" only on the PAMS test, which was given second in sequence.

Results

No effects due to sequence of test administration were found, nor were any differences due to sex. All groups were homogeneous with regard to ACE scores.

The results were strongly bimodal in the regular "one-block" administration of each test, scores of 1 through 5 having percentage values of 6, 32, 13, 11, and 38, respectively. This bimodality necessitates the use of chi square in evaluating correlations and differences between correlations.

TABLE 1

Correlations of Scores, Cond. 1 and 2

Scores on Ladd Test	Scores on PAMS Test		χ^2	*Phi*
	Above Median	Below Median		
Condition 1				
Above median	27	2		
Below median	6	32	39.3	.77
Condition 2				
Above median	10	8		
Below median	10	12	.44	.10

TABLE 2

Comparison of Proficiency Correlations, Cond. 1 and 2

	Condition	
	1	2
Above median both tests	27	10
Below median both tests	32	12
Above med. on Test 1 and below med. on Test 2	8	18

Comparison of Cond. 1 and 2.—When the distribution of scores in Cond. 1 and 2 are divided as closely to the medians as possible and are arranged in contingency tables (see Table 1), the correlation of scores in Cond. 1 differs significantly from zero ($\chi^2 = 39.3$, 1 *df*, $P < .001$), and that in Cond. 2 does not ($\chi^2 = .44$, 1 *df*, $P > .50$). The homogeneity of the relationship between the two conditions may be evaluated by computing χ^2 on the data of Table 2. Here it can be seen that the relationship between the test scores for the two groups differs significantly from zero ($\chi^2 = 15$, 2 *df*, $P < .001$).

Control of initial ability is furnished by comparison of performance on the PAMS test, which was given to both groups with the same administration. The means of the PAMS scores did not diverge significantly (Table 3). Thus the principal prediction of the theory presented above cannot be rejected. Administering both tests in a manner calculated to minimize inhibitory perseveration yields a high correlation between performance on the problems; when one test is administered in a manner to maximize perseveration, performance correlates

TABLE 3

COMPARISON OF CONDITIONS FOR INITIAL ABILITY BY IDENTICALLY ADMINISTERED TESTS

Test	Cond.	Mean	*SD*	*P*
PAMS	1	3.33	1.40	—
PAMS	2	3.42	1.48	
Ladd	1	3.36	1.37	—
Ladd	3	3.20	.89	

negligibly. The difference between the two correlations is such that it could not have occurred by chance.

Comparison of Cond. 1 and 3.— In Cond. 3 the correlation of proficiency on the two tasks was compared as in Table 1, and $\chi^2 = .16$ for 1 *df* ($P > .50$). Comparing the correlation with that of Cond. 1 (as in Table 2) a χ^2 of 14.5 was obtained (2 *df*, $P < .001$). The results thus indicate that requiring *S* to verbalize as he solves one of the problems significantly lowers proficiency correlation.

Ladd test performance in Cond. 1 and 3 was homogeneous, showing that *Ss* were equivalent in initial ability (Table 3). Condition 3 *Ss* took longer in solving the PAMS test than did Cond. 1 *Ss*, as might be expected, since thinking aloud necessarily takes time.

Comparison of Cond. 1 and 4.— In Cond. 4 *Ss* were informed they were timed. If this instruction had any demonstrable effect these *Ss* should take less time than the Cond. 1 *Ss* in performance on the tasks. This was correct: the untimed *Ss* averaged 17 min. for both tasks ($SD = 7.5$) and the timed *Ss* averaged 15 min. ($SD = 4.2$). The difference is not significant. This small difference is undoubtedly due in part to the fact that fewer timed than untimed *Ss* presented insightful solutions and the insightful solution correlates with shorter time.

The correlation of Ladd and PAMS performance is lower in Cond. 4 *Ss* than with Cond. 1 *Ss*, but the difference is not significant. However, a significant difference does appear with regard to proficiency. Pooling the two test scores for each *S* and comparing the resultant scores shows the timed *Ss* less proficient ($P = .02$).

In the untimed group (Cond. 1) the more proficient *Ss* took less time and the less proficient took more time. *Phi* for this relationship was .52; for the timed group (Cond. 4) the comparable *phi* was only .10. The performance of the untimed *Ss* may be characterized as realistic, for more time was used when greater difficulty was encountered. The timed (Cond. 4) *Ss*' performance was comparatively unrealistic, for time spent on the task had little relationship to the difficulty encountered.

These results seem to indicate behavioral disorganization in the timed group—lower proficiency is found; the proficiency correlation between the tasks decreases, and *Ss* do not spend time on the tasks commensurate with the difficulty encountered. But behavioral observations made during testing do not bear out this interpretation. While no formalized measure of tension was made, the timed *Ss* seemed to evince no more tension than the untimed *Ss*.

DISCUSSION

Two similar problems showed high correlation providing (*a*) frequent and optimally distributed clues were given during the administration, (*b*) *S* was not required to verbalize his thoughts, and (*c*) *S* was assured the tasks were not timed. Each of these conditions was systematically varied with each varia-

tion provoking an appreciable drop in correlation between the two tests, two of these three drops being highly significant. The timing instruction, which did not cause a significant drop in correlation, produced a significant drop in proficiency. The procedures used in Cond. 1 may or may not be necessary for attaining substantial correlations between reasoning problems. But no experiments to date have reported comparably high correlations.

It has been shown in many previous experiments that verbalization does not faithfully reflect thought process (**8**). The present work shows that required verbalization significantly alters thinking proficiency. Thus verbalization is not only an inadequate index of thought as Leeper indicates; it also disturbs the thought process, compared to what it is when verbalization is not required.

Maier (**9**) proposed several criteria which differentiate reasoning from learning. Our results suggest that susceptibility to time pressure is a possible additional characteristic of reasoning. Indeed, the present results and similar recent findings of Bruner, Goodnow, and Austin (**3**) strongly suggest an aspect of reasoning not heretofore recognized in the literature: at the high difficulty level with which we are concerned, reasoning seems to be an appallingly fragile process. It is manifested in a stable and predictable fashion in a short experimental session only when narrowly circumscribed conditions are met, and is manifested erratically when the conditions are slightly varied. It is not difficult to understand why few correlation coefficients of reasoning proficiency appear in the literature.

Summary

A theory of reasoning proficiency is advanced, and conditions are hypothesized which would make for high proficiency correlations between two problems. The correlation was found as predicted. Three other conditions were found to lower the correlation appreciably—holding back knowledge of results beyond an optimal frequency on one task, having *S* think aloud on one task, and telling *S* the performance was timed. With regard to the substantial correlation of difficult reasoning problems—here reported for the first time—it was suggested that the correlations' vulnerability to slight modifications in experimental conditions might be further explored as a distinguishing characteristic of reasoning when its object is a problem of considerable difficulty.

References

1. Adamson, R. E. Functional fixedness as related to problem solving: a repetition of three experiments. *J. exp. Psychol.*, 1952, **44,** 288–291.
2. Brunk, L. The one-block administration of Vigotsky-like tests. Mimeo., Univer. of Kansas, 1955.
3. Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.
4. Bulbrook, M. E. An experimental inquiry into the existence and nature of "insight." *Amer. J. Psychol.*, 1932, **44,** 409–453.
5. Claparède, E. La genèse de l'hypothèse. *Arch. de Psychol.*, 1933, **24,** 1–155.
6. Duncker, K. On problem solving. *Psychol. Monogr.*, 1945, **58,** No. 5 (Whole No. 270).
7. Guetzkow, H. An analysis of the operation of set in problem-solving behavior. *J. gen. Psychol.*, 1951, **45,** 219–244.
8. Leeper, R. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Chap. 19.
9. Maier, N. R. F. The behavior mechanisms concerned with problem-solving. *Psychol. Rev.*, 1940, **47,** 43–58.
10. Maier, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10,** 115–143.
11. Ray, W. S. Complex tasks for use in human problem-solving research. *Psychol. Bull.*, 1955, **52,** 134–149.
12. Reid, J. W. An experimental study of "analysis of the goal" in problem solving. *J. gen. Psychol.*, 1951, **44,** 51–69.
13. Selz, O. *Zur Psychologie des produktiven Denkens und des Irrtums.* Bonn: Cohen, 1922.
14. Wertheimer, M. *Productive thinking.* New York: Harper, 1945.

(Received February 19, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

EYELID CONDITIONING AS A FUNCTION OF UNCONDITIONED STIMULUS INTENSITY AND INTERTRIAL INTERVAL [1]

WILLIAM F. PROKASY, JR., DAVID A. GRANT, AND NANCY A. MYERS

University of Wisconsin

It is well established that lengthening the intertrial interval (**1, 3, 9**) and increasing the UCS intensity (**5, 7, 8, 10**) will independently increase the level of eyelid conditioning. Unpublished research at the Wisconsin laboratory, however, suggests that high levels of the UCS combined with short intertrial intervals produce less conditioning than a UCS of lower intensity. In the light of these results it was thought that complicated interacting factors might be operating in the eyelid conditioning situation. More specifically, it was thought that high UCS intensity, when combined with fairly short intertrial intervals might result in sensory adaptation of the cornea, which would reduce the effective intensity of the UCS. Therefore, it was the purpose of this study to investigate the effects of UCS intensity, intertrial interval, and their interaction on the acquisition and extinction of the conditioned eyelid response. If the above conjecture is correct, the intensity-interval interaction will be significant, with the superiority of the high intensities most marked in procedures which also incorporate a fairly lengthy intertrial interval.

METHOD

Apparatus.—The apparatus was a modification of the one employed by Grant, Schipper, and Ross (**3**). The essential change was that the recording equipment incorporated differentiators which permitted accurate determination of response latencies. The signal from the microtorque potentiometer at the eyelid was amplified and differentiated so as to give a recording of the eyelid movement and its first derivative by means of an inkwriter system.

The CS consisted of a light which lasted 750 msec. and the UCS was a corneal air puff administered 500 msec. after the onset of the light. The *Ss* were run under standard light adapted conditions to minimize the presence of sensitized beta responses (**2**). The puff was generated by the fall of a column of mercury in a manometer, and was applied to the left eye. Eyelid movements were recorded from the right eye.

Procedure.—All *Ss* were given 20 training trials on Day 1 of the experiment. On Day 2, 20 training trials, followed by 20 extinction trials, were administered. There were three average intertrial intervals employed during acquisition: 15, 45, and 135 sec. The average intertrial interval was 45 sec. during extinction for all *Ss*. All intervals were varied randomly within 5 sec. of their average. There were four levels of UCS intensity, as gauged by the fall of a column of mercury: 50, 120, 190, and 260 mm. This resulted in a 3×4 orthogonal design with 10 *Ss* in each cell. A CR was defined as any response that occurred between 250 and 500 msec. after the onset of the CS.

Subjects.—The *Ss* were 120 men and women students from introductory classes in psychology at the University of Wisconsin. They were assigned at random to the 12 experimental groups employed. They were instructed to maintain a passive attitude toward the experiment and not to influence their natural reactions to the stimuli.

RESULTS

In Fig. 1 is plotted the percentage frequency of CR's for each five-trial block during training and extinction for the three intertrial intervals. Figure 1 thus illustrates the over-all

[1] This research was supported in part by the Research Committee of the Graduate School of the University of Wisconsin with funds from the Wisconsin Alumni Research Foundation.

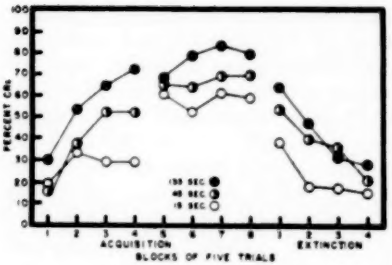

FIG. 1. Percentage frequency of CR's during acquisition and extinction trials with intertrial interval during acquisition as the parameter.

effects of intertrial interval. The results here are quite conventional, being in line with those of Baron (**1**), Spence and Norris (**9**), and Grant, Schipper, and Ross (**3**). The longest intertrial interval is associated with superior conditioning and greater resistance to extinction, this effect being significant for Day 1 and Day 2 acquisition and extinction, by the analyses of variance of the over-all scores in Table 1. Considerable overnight recovery is shown for the two relatively massed groups, suggesting that a decremental process such as reactive inhibition developed during Day 1 and dissipated overnight.

In Fig. 2 are plotted percentages of CR's during acquisition and extinction with UCS intensity as the parameter. Here, again, the results are essentially conventional in that superior conditioning was obtained overall with the most intense puff, and the least percentage of CR's was found with 50-mm. puff. The effects of intensity were significant on both Day 1 and Day 2 acquisition, but not significant during extinction, as shown in Table 1. It is apparent that the significant intensity effect of Day 2 is primarily due to the low conditioning level of the 50-mm. puff groups, since overnight recovery of the middle two intensity groups, though slight, is sufficient to eliminate differences between the three highest intensity groups. Only slight overnight recovery in Fig. 2 should be expected if that demonstrated in Fig. 1 is due to a reactive inhibitory process, since increasing the UCS intensity should not produce as much cumulative inhibition as massing trials.

Figure 3 depicts the level of conditioning as a function of intensity with intertrial interval as parameter for Day 1 and Day 2 acquisition. In general, as intensity increases, so does conditioning. However, a significant interaction failed to materialize on Day 1 (see Table 1), although the increase in conditioning from 190 to 260 mm. was greater for the longest

TABLE 1

ANALYSIS OF VARIANCE OF DAY 1 AND DAY 2 ACQUISITION AND EXTINCTION
(arc sine transformation of percentage frequency)

Source	*df*	Day 1—Acq. *F*	Day 2—Acq. *F*	Day 2—Ext. *F*
Intertrial interval	2	7.47**	4.72**	5.16**
UCS intensity	3	7.51**	9.14**	2.19
Interaction	6	1.34	2.29*	1.23
Error	108	(253.18)	(427.13)	(367.02)

Note.—Mean squares for error in parentheses.

* $P = .05$.
** $P = .01$.

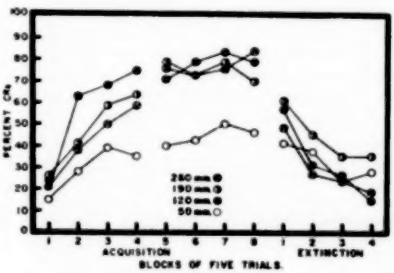

FIG. 2. Percentage frequency of CR's during acquisition and extinction with UCS intensity during acquisition as the parameter.

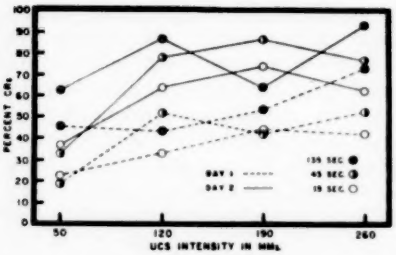

FIG. 3. Over-all percentage CR's during Day 1 and Day 2 acquisition trials as a function of UCS intensity with intertrial interval as the parameter.

intertrial interval, as predicted. The Day 2 acquisition interaction was significant, but this was due primarily to the aberrant point (presumably a sampling error) at the 190-mm. puff level and 135-sec. intertrial interval. The level of conditioning drops for the two relatively massed intertrial intervals at the highest intensity, which might be expected if there is sensory adaptation of the cornea at very high intensities and short intervals. The same plot for extinction data (not shown) reveals essentially that the percentage of response is the same for all groups, regardless of the UCS intensity on the preceding acquisition trials.

Table 2 presents an analysis of covariance, with the Day 2 acquisition data adjusted for differences in Day 1 responding, and with the extinction data adjusted for the response level on Day 1 and Day 2 acquisition combined. After covariance adjustment, only the UCS intensity effects remained significant during Day 2 acquisition, indicating that the interval effect occurs on Day 1, and is continued, but not increased disproportionately, on Day 2. Intensity of the UCS, on the other hand, has a continuing and increasing effect on Day 2 which cannot be accounted for entirely in terms of linear regression from Day 1 acquisition.

TABLE 2

ANALYSIS OF COVARIANCE OF DAY 2 ACQUISITION AND EXTINCTION (arc sine transformation of percentage frequency)

Source	*df*	Acquisition *F*	Extinction *F*
Intertrial interval	2	1.0	2.23
UCS intensity	3	3.48*	2.72*
Interaction	6	1.78	1.0
Error	107	(314.20)	(170.96)

Note.—Mean squares for error in parentheses.
* $P = .05$.

Although only the intertrial interval effect was significant in the analysis of the raw extinction data, controlling for response level during acquisition by analysis of covariance eliminated this difference, and revealed a significant intensity effect. It was evident from this analysis that differences in performance caused by UCS intensity during acquisition did not permit prediction of performance during extinction. The disappearance of significant interval effects after covariance adjustment suggests that differences in extinction due to prior differences in intertrial interval (since intervals for all groups were the same during extinction) are a continuation of the over-all differences produced during acquisition, with nothing new appearing during extinction.

DISCUSSION

Aside from the conventional results on the effect of UCS intensity and interval upon the acquisition of the conditioned eyelid response, the present experiment and subsequent covariance analyses give an indication of where these variables operate during the conditioning process. Despite slight overnight recovery, it appears that intertrial

interval differences have an immediate and constant effect during the first 20 trials, which, when eliminated by covariance analysis produce no further differences during the second 20 trials of the experiment. Baron (1) has suggested that, asymptotically, conditioning is a negatively accelerated increasing function of intertrial interval. Plotting the present data in this form (not shown) suggests more of a linear relationship within the range tested; but this is by no means clear cut. The difference between Baron's suggestion and the present findings may be due to the 24-hr. break between the two 20-trial acquisition sessions. The possible distortion due to overnight recovery remains to be investigated more extensively.

Covariance adjustment also brought out a significant difference in extinction as a function of prior UCS intensity. This can be explained at least partially in terms of the differences between the acquisition and extinction stimulus situations. Even if the amount of conditioning were higher for the high intensity groups, removal of the UCS during extinction might constitute a more drastic change for those groups and hence lead to relatively faster extinction. However, the fact that the differences in unadjusted extinction scores do not turn out to be significant may not necessitate this interpretation. A phenomenon similar to that in certain instrumental reward studies with rats (4, 6, 11) may be operating. In these experiments, increased resistance to extinction was not associated with higher drives during acquisition.

There is some indication (Fig. 3) that the heightening of conditioning due to higher levels of the UCS is more readily obtained when the intertrial interval is lengthened. Figure 3 suggests, in fact, that there is a drop in the level of conditioning for the highest UCS intensity and the shortest intervals. For the longest interval, conclusions are obscured because of the aberrant point which made the primary contribution to the significant interaction in Day 2 acquisition. Passey (5) concluded that level of conditioning is a negatively accelerated, increasing function of intensity. The data for the first 20 conditioning trials do not support this contention, but support is found in the final 20 conditioning trials (Fig. 3). The approach to the asymptote is probably a function of the intertrial interval.

Grant, Schipper, and Ross (3) demonstrated that massed extinction following spaced acquisition is faster than spaced extinction following spaced acquisition, and, further, that spaced extinction following massed acquisition is faster than massed extinction following spaced acquisition. This phenomenon appears in the extinction curves of Fig. 1. All groups were switched to 45-sec. intertrial intervals during extinction, which would suggest that relative to the 45-sec. acquisition group the other two groups would extinguish more rapidly. This proved to be the case, as the 15-sec. interval group extinguished most rapidly, while the 135 and 45-sec. group curves nearly overlap during extinction in spite of the fact that the 135-sec. interval group was definitely superior during acquisition.

Summary

An experiment was designed to investigate the effects of four levels of UCS intensity and three levels of intertrial interval upon the acquisition and extinction of conditioned eyelid responses ($N = 120$). In general, as UCS intensity increased and as intertrial interval lengthened, percentage frequency of CR's also increased. There was some suggestion of an interaction with high intensities and short intervals resulting in a slightly lower level of responding.

The effects of intertrial interval were immediate and essentially complete during the first 20 conditioning trials, with no further differences resulting during extinction. UCS intensity effects were significant during both the first and second sets of 20 acquisition trials, and, when means were adjusted by means of covariance, were also significant during extinction.

References

1. Baron, M. R. The effect of long intertrial intervals on the limit of eyelid conditioning. *J. exp. Psychol.*, 1952, **44**, 438–441.

2. GRANT, D. A., Norris, E., & BOISSARD, S. Dark adaptation and the pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1947, **37**, 434–439.
3. GRANT, D. A., SCHIPPER, L. M., & ROSS, B. M. Effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.
4. KENDLER, H. Drive interaction: II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, 1945, **35**, 188–198.
5. PASSEY, G. E. The influence of intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.
6. REYNOLDS, B. The relationship between strength of habit and the degree of drive present during acquisition. *J. exp. Psychol.*, 1949, **39**, 296–305.
7. RUNQUIST, W. N., SPENCE, K. W., & STUBBS, D. W. Differential conditioning and intensity of the UCS. *J. exp. Psychol.*, 1958, **55**, 51–55.
8. SPENCE, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
9. SPENCE, K. W., & NORRIS, E. B. Eyelid conditioning as a function of intertrial interval. *J. exp. Psychol.*, 1950, **40**, 305–313.
10. SPENCE, K. W., & TAYLOR, J. Anxiety and strength of UCS as determinants of amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
11. STRASSBURGER, R. C. Resistance to extinction of a conditioned operant as related to drive level at reinforcement. *J. exp. Psychol.*, 1950, **40**, 473–487.

(Received March 14, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

VERBAL AND MOTOR RESPONSES TO SEVEN SYMBOLIC VISUAL CODES: A STUDY IN S-R COMPATIBILITY [1]

EARL A. ALLUISI [2] AND PAUL F. MULLER, JR.[3]

Laboratory of Aviation Psychology, The Ohio State University

Man's ability to handle information depends not only upon the specific stimuli and responses he uses, but also upon their combination in the task he performs. The dependence of performance on this interaction has been called "S-R compatibility" (**4; 5; 13,** pp. 316–341).

Although S-R compatibility effects have been illustrated in several experimental studies (**2, 7, 8, 9, 10, 12**), it is not always possible to predict which stimulus and response conditions will prove to be the most compatible of a series of S-R sets. It does appear, however, that relative S-R compatibility tends to approach a maximum when (*a*) the stimulus and response sets correspond to one another in a direct physical sense, and when (*b*) the pairings of stimulus and response elements agree with strong population stereotypes, where such stereotypes exist (**13,** p. 320).

On the basis of the population stereotype, it might be predicted that when verbal responses of a numerical nature are to be made, the most compatible set of symbols are those of the conventional Arabic numerals. Men have been well overtrained in recognizing and interpreting these symbols as ordinal scale values and in responding to them as "one," "two," etc.

There is no apparent way to predict relative S-R compatibility when other stimuli are used in the same way, even when these symbols form easily discriminated ordered series. What data are available indicate only that the information-handling performance obtained with conventional Arabic numerals cannot be surpassed with other symbolic coding schemes (**3**).

The present study was specifically designed to measure information-handling performance (**14**) of *S*s responding to conventional Arabic numerals and to six other symbolic visual codes. The conventional Arabic numerals were used for purposes of comparison; the other sets of symbols are alternative, and technically feasible, ways of encoding information for display on cathode ray tubes and other electronic devices. Motor (key-pressing) and verbal responses were both used in order to lend greater generality to the results and to extend the study of S-R compatibility effects.

[1] This research was supported in part by the U. S. Air Force under Contract No. AF 33(616)-43, Project No. 7192, with the Ohio State University Research Foundation, monitored by the Aero Medical Laboratory. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government. The authors wish to acknowledge the assistance of Hugh B. Martin in the operation of the apparatus, of Ilse B. Webb in the analysis of the data, and of both Paul M. Fitts and Ralph Queal who made many helpful contributions during the entire study.

[2] Now at the Experimental Psychology Department, U. S. Army Medical Research Laboratory, Fort Knox, Kentucky.

[3] Now at the Missile Systems Division, Lockheed Aircraft Corporation, Sunnyvale, California.

METHOD

Two experiments were conducted; they differed essentially in requiring different types of responses. In Exp. I, *Ss* responded by pressing one of a number of finger keys for each of the symbols presented. In Exp. II, *Ss* responded verbally by calling out the number assigned to each stimulus symbol. A rest interval of four weeks intervened between the two experiments; during that time, *Ss* served in no experiments.

Apparatus.—The OSU Serial Discrimeter was the apparatus used in this study; this has been described in greater detail elsewhere (**1, 10**), and is similar in function to the Multiple Serial Discrimeter described by Morin and Grant (**9**). It provided a means of serially presenting single stimuli from a 10-stimulus set in any desired sequence at self-paced, or at a wide variety of preselected forced-paced rates.

A bank of 10 finger keys was used by *Ss* to respond in Exp. I. The keys were arranged horizontally in two semicircles to correspond with the natural placement of the 10 fingers. The keys were numbered serially (*1*, *2*, etc.) from left to right, and *S* was told to respond by pressing the key corresponding to the symbol presented. In Exp. II, *Ss* responded by speaking into a boom microphone. During both experiments, *Ss* were presented approximately 70 db of broad-band noise through earphones to mask extraneous sounds.

An automatic scoring unit recorded all stimulus and response signals on a matrix of counters during Exp. I. The automatic recording matrix was not used during Exp. II because the scoring unit could not discriminate among the different verbal responses; instead, error patterns were recorded by a monitor using lists of the programmed stimuli, and tape recordings were used to resolve any scoring ambiguities. A Standard Electric timer was used to record the total time for a trial of 100 stimulus presentations in both experiments.

Because of the lag of responses behind stimuli, scoring delays of .10, .20, and .30 sec. were used with the forced-paced rates of 4, 5, and 6 bits/sec., respectively, when the automatic recording matrix was used. These delays had been previously determined to be optimal for the stimulus presentation rates employed. No scoring delay was necessary under the other conditions of the experiments.

Stimuli.—The symbols used as stimuli in this study are represented in Fig. 1. One of the two numerical codes used was the standard AND 10400 numerals; the other was a symbolic alphabet of straight-line Arabic numerals adapted (**11,** pp. 84–89) from those initially proposed by Cohen and Webb (**3**).

The colored circles were produced by use of gelatin filters as described elsewhere (**11,** p. 13). The colors employed could be described as dark (D) and light (L) shades of blue (B), green (G), and red (R), and single shades of gray (Gy), yellow (Y), orange (O), and violet (V). Psychophysical data relating to the other symbols, including equal discriminability scales for the coding dimensions of inclination and ellipse-axis ratio, have been previously reported (**11**).

The stimuli were projected from photographic high-contrast negatives onto the rear of a 10-in. diameter opal glass screen by means of a 10-unit optical projection system; they appeared as light patterns at the center of the display screen approximately 28 in. in front of *S*. The dimensions of the code symbols were proportional to those shown in Fig. 1, with the height of the symbolic Arabic numerals set at .25 in. The experimental booth was dimly illuminated, brightness contrast was relatively high, and all the stimuli were quite legible.

Subjects.—Ten male *Ss* were used in Exp. I. They were obtained from a pool of students who had volunteered to serve for pay in laboratory experiments. Nine of these *Ss* served in Exp. II. The tenth *S* used in Exp. II was a member of the laboratory staff who replaced one *S* lost from the original group. This last *S* was given training equivalent to that received by the other *Ss* in Exp. I.

Procedure.—The 10 *Ss* were divided into pairs before each of the experiments. Each pair served for an hour-long experimental session on each day of experimentation. During each session, *S* responded for one trial of 100 stimulus presentations with each of the seven codes.

The 100 stimuli in each trial were arranged in a different random order with each of the ten stimulus categories appearing equally often. The orders of presentations of trials (codes) were also random, within the restriction that each code was represented once and only once in each block of seven trials (a half session).

In a given session, the two *Ss* in a pair were seated in the experimental booth. One *S* faced the projection screen directly, and the other *S* sat to his left somewhat off the direct line of sight to the screen. The first *S* was in the "operator" position, and the second *S* was in the "nonoperator" position. During Exp. I, the operator-*S* responded with the primary bank of finger keys, while the nonoperator-*S* practiced with an identical bank of dummy finger keys. During Exp. II, the operator-*S* responded by speaking into the microphone, while the nonoperator-*S* recorded response errors on a list of the programmed stimuli.

During both experiments, after the first *S*

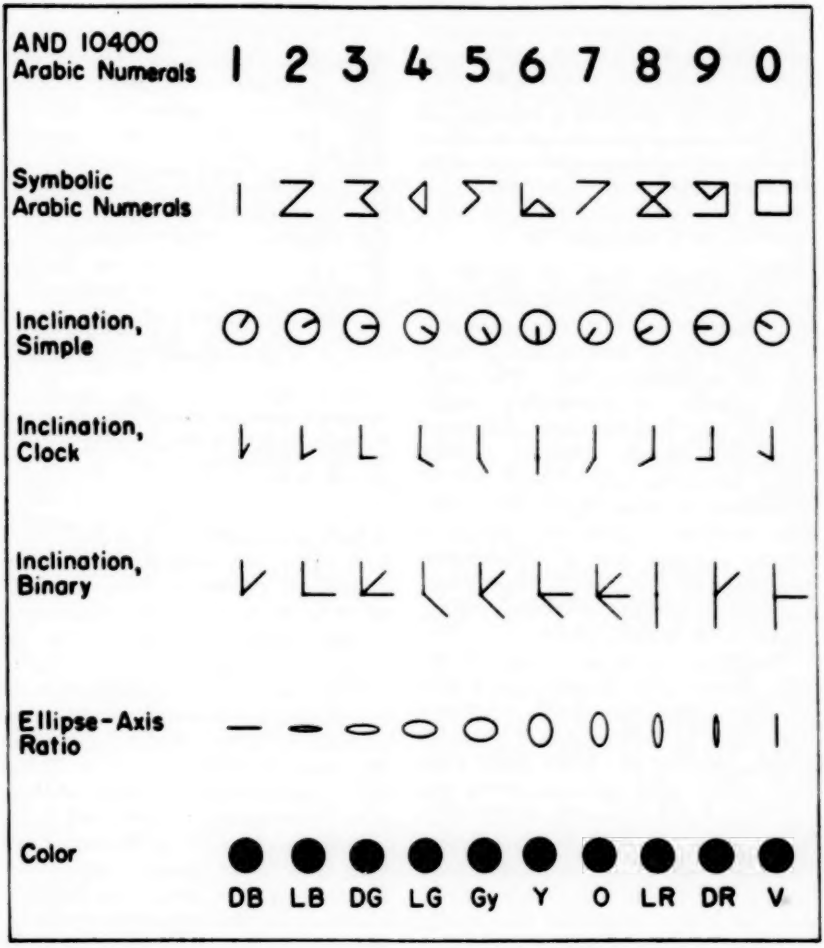

FIG. 1. The seven symbolic visual codes used in this study. The columns from left to right represent the numerals from *1* to *10*.

had responded for seven trials, *S*s changed positions and the second *S* assumed the operator position. The operator-nonoperator positions were balanced within pairs of *S*s in an ABBA order over the experimental sessions so that *S* was equally often in the operator position during the two half sessions.

Experiment I consisted of 16 sessions per *S*, one session on each day of the experiment. These sessions were divided into three sections: six days of self-paced pretraining were followed by five days of forced-paced performance, and these were followed by a final five days of self-paced posttraining performance. The sessions of Exp. II were similar: four days of self-paced pretraining trials were followed by five days of forced pacing, and these were followed by a final three days of self-paced posttraining performance.

In both experiments, the rates of information presentation (H_s, in bits/sec.) used in the forced-paced trials varied from 2 to 6 bits/sec. in unit

steps. These rates were obtained by use of five rates of stimulus presentation: one stimulus every 1.661, 1.107, .831, .664, and .554 sec. for rates of 2, 3, 4, 5, and 6 bits/sec., respectively. All five rates were used on each day of the forced-paced trials with each pair of *Ss* responding to a single, but different rate. Over the five days of forced pacing, each *S* responded to each of the five rates with each of the seven codes.

Results

The response data of all 10 *Ss* were pooled into 196 S-R matrices—one matrix for the 1000 stimulus presentations of each code on each day of experimentation. The average amount of information transmitted per *S* was then computed from each matrix. Finally, each of these estimates was divided by its associated average responding time; this resulted in producing estimates of the average rate of information transmission (H_t, in bits/sec.) obtained with each of the seven symbolic visual codes in each session.

Friedman's chi-square-*r* statistic (**6**) was applied as a nonparametric test of the significance of the obtained differences in performance. The differences among the seven codes were first tested with the sessions of the two experiments taken as replications. The results indicated statistically significant differences in performance among the codes over both the 16 days of motor responses in Exp. I, and the 12 days of verbal responses in Exp. II ($P < .001$, in each case).

In both experiments the codes fell into three nonoverlapping groups: the two numerical codes (i.e., conventional and symbolic Arabic numerals), the three inclination codes, and the two remaining codes of color and ellipse-axis ratio. Differences within these groups appeared to approach statistical significance only with the color and ellipse-axis codes in Exp. I, and with the numerical codes

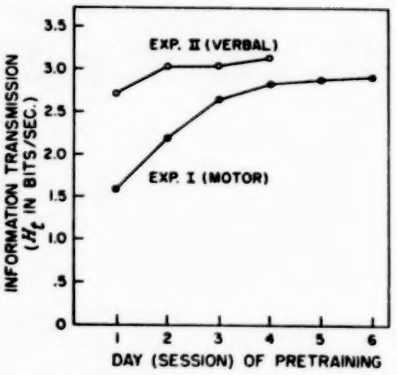

Fig. 2. Rate of information transmission during the self-paced pretraining trials of each experiment. Each point is based on 1000 stimulus presentations of each of the seven symbolic codes.

in both experiments ($.02 < P < .05$, in each case); the magnitudes of these differences, however, were so small as to be of no practical importance, so they have been ignored in the remainder of this paper.

Pretraining trials.—A different number of days, or sessions, of self-paced pretraining trials was employed in the two experiments. The object was to train *Ss* to a criterion of three days of relatively stable self-paced performance before presenting them with the forced-paced conditions. The success with which this criterion was met in the six days of motor-response training and the four days of verbal-response training is illustrated in Fig. 2, where H_t has been further averaged over all seven symbolic codes.

More pretraining was required to reach the criterion of performance with motor responses than with verbal responses. This result was not entirely unexpected; each *S* experienced 1600 stimulus presentations of each code in Exp. I before he responded to the first of the pretraining

trials in Exp. II. It is reasonable to assume that this experience might help shorten the length of practice required to reach a relatively stable level of performance, even though different types of responses were involved.

Also, the pretraining trials of Exp. I involved the development of the specific motor skills necessary for making the key-pressing responses; such specific learning was minimized in Exp. II by use of the already highly developed skill for making verbal responses of "one," "two," etc.

Forced-paced performance.—The average rate of information transmission per *S* was computed for each of the codes at each of the five rates of information presentation. These data are illustrated in Fig. 3 for both experiments (for simplicity of presentation, H_t has been further averaged into the three groupings of numerical codes, inclination codes, and codes of color and ellipse-axis ratio).

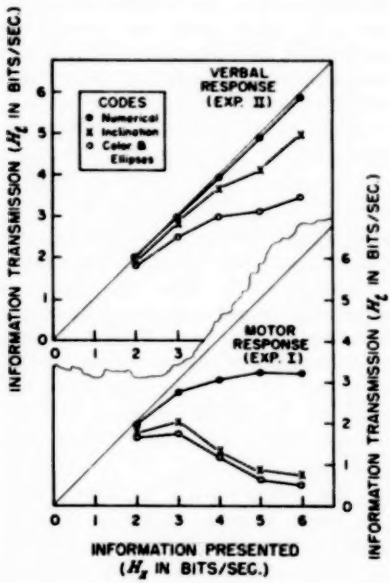

FIG. 3. Rate of information transmission as a function of the rate of information presentation. The solid diagonal line represents the maximum possible transmission rates.

The verbal-response data indicate relatively little overloading of the information-handling system with the conditions of forced pacing in Exp. II; i.e., over the five levels of H_s sampled in Exp. II, none of the information transfer functions reached a maximum and then fell off with further increases in H_s. The motor-response data, on the other hand, indicate that the information-handling ability of *S* was definitely overloaded with the nonnumerical codes at levels of H_s above 3 bits/sec. (with the two numerical codes, maximum performance was reached at or above the H_s of 4 bits/sec.).

Verbal-response performance with the two numerical codes was nearly perfect at all the values of H_s used here. However, with increases in H_s above the minimum rate of 2 bits/sec., equivocation quite obviously increased in the verbal responses to the nonnumerical codes, and in the motor responses to all codes. This is illustrated in Fig. 3 by the increases in the distances between the diagonal straight line (maximum performance possible) and the separate functions as H_s goes from 2 to 6 bits/sec. This means that omissions and errors in responding increased with each increase in H_s above the minimum rate employed.

Previously reported data (1) obtained with conventional Arabic numerals and the same types of responses used here indicate that the limit to *S*'s information-handling performance is not the rate of information presentation, per se, but rather the rate of stimulus presentation, and that equivocation increases with increases in

TABLE 1

AVERAGE POSTTRAINING SELF-PACED PERFORMANCE WITH SEVEN SYMBOLIC CODES

Code	Speed[a]	Accuracy[b]	H_t[c]
AND 10400	72.70	93.08	3.97
Arabic Numerals	*75.11*	*99.57*	*4.38*
Symbolic	73.22	93.96	4.00
Arabic Numerals	*77.91*	*97.00*	*4.20*
Simple	92.85	91.26	3.05
Inclination	*107.14*	*99.03*	*3.02*
Clock	90.96	92.82	3.17
Inclination	*103.33*	*98.67*	*3.14*
Binary	90.50	93.08	3.23
Inclination	*103.26*	*97.73*	*3.13*
Ellipse-Axis	100.27	88.86	2.68
Ratio	*122.99*	*99.37*	*2.56*
Color	100.05	88.94	2.74
	120.50	*98.87*	*2.60*

Note.—Upper values are data for the motor responses of Exp. I; lower italicized values are data for the verbal responses of Exp. II.

[a] Mean time, in sec., per 100 responses.

[b] Percentage correct responses.

[c] Average rate of information transmission per *S*, in bits/sec.

the stimulus presentation rate above 1 stimulus/sec. In general, the data of the present experiments are consistent with this previous finding.

Posttraining performance.—The forced-paced conditions in both experiments provided Ss additional practice in responding to the codes before the final measurements were made of their self-paced performance. For this reason, the two final series of self-paced trials have been called "posttraining" performances.

In the case of each code in both experiments, the mean posttraining performance was higher than the mean pretraining performance (the latter measured over the last three days of relatively stable pretraining). The average gain in information-handling performance was 13.5% with motor responses, and 6.9% with verbal responses. These gains appear sufficiently high to justify the apparent inconsistency of design in interpolating forced pacing between preliminary and final measurements of self-paced performance.

The actual mean rates of information transmission obtained during the posttraining trials are given in Table 1 along with the mean speed and accuracy data of the same series of self-paced trials. Each tabled value for motor responses is based upon 5000 stimulus presentations; each tabled value for verbal responses is based upon 3000 stimulus presentations.

The data of Table 1 illustrate several points. First, the greater accuracy of the verbal responses was about compensated for by the greater speed of the motor responses. This is illustrated in the information measure, H_t, that combines these two aspects of performance. This result is considerably different from that obtained with forced pacing.

Secondly, the data further illustrate the general performance groupings of the codes into (*a*) the two numerical codes, (*b*) the three inclination codes, and (*c*) the codes of color and ellipse-axis ratio. Differences between codes within any of these groups were relatively small, whereas differences between codes in different groups were relatively large. This is interpreted as an indication that the different stimulus dimensions may be generally reflecting different relative task difficulties as represented by the performances obtained with the three groups. That is to say, inclinations may require more time to be recognized and interpreted as ordinal scale values than numerals, and colors or ellipse-axis ratios may require still more time than inclinations.

Finally, the data of Table 1 il-

lustrate the not surprising superiority of the conventional Arabic numerals as stimuli for ordering responses of either a verbal or motor kind. The closeness with which this performance was approached by use of the "simplified" set of straight-line symbolic Arabic numerals was not entirely expected, however, because other data have suggested that greater differences in performance might exist (**3**). Resolution of this difference must await the collection of additional empirical evidence.

Discussion

The usefulness of the average rate of information transmission (H_t) as a measure of performance is well illustrated in the data of Table 1. Obviously, if accuracy had been taken as the sole criterion of this self-paced performance, verbal responses would have appeared superior to motor responses in the case of each of the seven codes. On the other hand, if only speed had been taken, motor responses would have appeared superior to verbal. These are indeed the results if one wishes specifically to stress speed or accuracy to the point of ignoring the other criterion.

However, both speed and accuracy may be combined into the single information measure, H_t, on the basis of the theory provided by Shannon and Weaver (**14**). The advantage of a combined speed-accuracy score lies in the possibility of reaching general conclusions regarding performance without specifically stressing either speed or accuracy. The advantage of using H_t as this combined score lies in the well developed logic of information theory that at least theoretically justifies the specific weightings of speed and accuracy that are intrinsic to the combining of these scores. In this it is believed that H_t serves as a useful metric for psychological research.

Further inspection of the data of Table 1 will reveal an interesting interaction in the H_t data: H_t with verbal responses was apparently slightly higher than with motor responses for the two highest (numerical) codes, it was about equal for the intermediate (inclination) codes, and it was slightly lower for the two lowest codes. This interaction may be interpreted as an illustration of S-R compatibility effects.

Still another illustration may be taken from the forced-paced data of Fig. 3. The type of stimulus apparently had an effect upon performance that was independent of the mode of response; e.g., with both motor and verbal responses the numerical codes were superior to the inclination codes, and these were superior to the codes of color and ellipse-axis ratio. Also, the mode of response apparently had an effect upon performance that was independent of the stimulus; e.g., the best motor-response performance was about equal to the worst verbal-response performance. Finally, stimulus code and response mode apparently interacted as shown by the relative similarity of the motor-response performances with the non-numerical codes, and the relative dissimilarity of the verbal-response performances with the same codes.

In other words, performance is determined by both the stimulus and the response conditions, not independently, but in the dependent way implied in the concept of S-R compatibility (cf. **13**, pp. 316–341). It is the interaction of stimuli with responses that gives rise to S-R compatibility effects.

Summary

The information-handling performance of 10 *Ss* was measured with conventional Arabic numerals and six other symbolic visual codes under both motor- (key-pressing) and verbal-response conditions. The six symbolic codes included a set of straight-line symbolic Arabic numerals, three sets of ordered symbols based upon differences in the visual inclination of a line, a set of colors, and a set of ellipses of differing axis ratios. The two response modes were used to lend generality to the results and to extend the study of S-R compatibility effects. Both self- and forced-paced rates of information presentation were used, with the latter being

varied from 2 to 6 bits/sec. in unit steps. The major results were as follows:

1. The seven codes fell into three clearly different groups under each response condition: the two numerical codes were superior to the three inclination codes, and these were all superior to the two remaining codes of color and ellipse-axis ratio.

2. Verbal-response performance with the two numerical codes was nearly perfect over the forced-paced rates used here. However, under all other forced-paced conditions, equivocation increased with increases in the rate of information presentation above the minimum of 2 bits/sec.

3. In self pacing, verbal responses with each code were made with greater accuracy than motor responses, but motor responses were made with greater speed than verbal.

4. Interactions of stimulus codes with response modes were found in both self- and forced-paced performances; these interactions were interpreted as illustrations of S-R compatibility effects.

References

1. Alluisi, E. A., Muller, P. F., Jr., & Fitts, P. M. An information analysis of verbal and motor responses in a forced-paced serial task. *J. exp. Psychol.*, 1957, **53,** 153–158.
2. Anderson, N. H., Grant, D. A., & Nystrom, C. O. Performance on a repetitive key pressing task as a function of the spatial positioning of the stimulus and response components. *USAF, WADC Tech. Rep.*, 1954, No. 54-76.
3. Cohen, J., & Webb, I. B. An experiment on the coding of numerals for tape presentation. *USAF, WADC Tech. Rep.*, 1953, No. 54-86.
4. Fitts, P. M., & Deininger, R. L. S-R compatibility: correspondence among paired elements within stimulus and response codes. *J. exp. Psychol.*, 1954, **48,** 483–492.
5. Fitts, P. M., & Seeger, C. M. S-R compatibility: spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46,** 199–210.
6. Friedman, M. The use of ranks to avoid the assumptions of normality implicit in analysis of variance. *J. Amer. statis. Ass.*, 1937, **32,** 675–701.
7. Garvey, W. D., & Knowles, W. B. Response time patterns associated with various display-control relationships. *J. exp. Psychol.*, 1954, **47,** 315–322.
8. Garvey, W. D., & Mitnick, L. L. Effect of additional spatial references on display-control efficiency. *J. exp. Psychol.*, 1955, **50,** 276–282.
9. Morin, R. E., & Grant, D. A. Learning and performance on a key-pressing task as a function of the degree of spatial stimulus-response correspondence. *J. exp. Psychol.*, 1955, **49,** 39–47.
10. Muller, P. F., Jr. Verbalization as a factor in verbal vs. motor responses to visual stimuli. Unpublished doctoral dissertation, Ohio State Univer., 1955.
11. Muller, P. F., Jr., Sidorsky, R. C., Slivinske, A. J., Alluisi, E. A., & Fitts, P. M. The symbolic coding of information on cathode ray tubes and similar displays. *USAF, WADC Tech. Rep.*, 1955, No. 55-375.
12. Nystrom, C. O., & Grant, D. A. Performance on a key pressing task as a function of the angular correspondence between stimulus and response elements. *USAF, WADC Tech. Rep.*, 1954, No. 54-71.
13. Quastler, H. (Ed.) *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955.
14. Shannon, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana: Univer. Ill. Press, 1949.

(Received March 26, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

"PERMANENT" SATIATION PHENOMENA WITH KINESTHETIC FIGURAL AFTEREFFECTS[1]

MICHAEL WERTHEIMER AND CARL M. LEVENTHAL[2]

University of Colorado

Prolonged inspection of an object in the Köhler-Wallach paradigm (**4**) produces figural aftereffects (FAE) which may persist for long periods of time. The longevity of these phenomena was first noted by Köhler and Wallach (**4**), and additional information has been provided by further studies of both visual and kinesthetic FAE (**1, 3, 4**). Any FAE is maximal immediately after its induction, and then dissipates slowly, with a negatively accelerated decay function which may never return to the original pre-inspection level, thus leaving a residual FAE which Köhler has called the "permanent" FAE, or "permanent satiation."

Although Köhler and Dinnerstein (**2**) provide some incidental observations on the persistence of the kinesthetic FAE, systematic studies of such persistence have not been reported. The aim of the present study was to measure the size of the residual kinesthetic FAE as a function of the amount of inspection.

METHOD

The apparatus and technique used in the measurement of the kinesthetic FAE were similar to those used by Köhler and Dinnerstein (**2**) and are described elsewhere (**5**). The *S* stood between two pairs of parallel bars. In front of *S* were the test bars—a 1½-in.-wide standard bar on the left and a bar on the right which was 1 in. wide at the near end and increased in width by a total of thirty $\frac{1}{16}$-in. steps to 2$\frac{7}{8}$ in. at the far end. Behind *S* were two "inspection" bars; the left bar was ¾ in. wide and the right bar 2¾ in. wide. All bars were 60 in. long and 2 in. high, and were 35 in. above the floor.

The blindfolded *S* was requested to find the place on the right-hand test bar which felt just as wide as the left-hand test bar. A test consisted of 4 successive trials; on Trials 1 and 4 *S* started on the variable at a width below that of the standard, and on Trials 2 and 3 *S* started on the variable at a width considerably greater than that of the standard. Following these four pre-inspection matches, *S* was asked to step back and rub the "inspection" bars for a certain period, and after this inspection, four post-inspection trials were made, using the identical sequence as in the pre-inspection trials. The size of the FAE was measured as the difference between the mean width chosen during the pre-inspection test and the mean width chosen during the post-inspection test. The size of the residual FAE was measured by the difference between the mean width chosen during the pre-inspection test on the first day of the study and the mean width chosen during the pre-inspection test on the next day of the study.

The apparatus was so arranged that an increase in mean test match was evidence of a FAE. Thus, if the mean pre-inspection test match on Day 2 were higher than the mean pre-inspection test match on Day 1, this would indicate a residual FAE (suggestive of "permanent" satiation) induced by the inspection on Day 1.

All *S*s came at the same time of day for five successive days. They were divided into three groups: Group 1 ($N = 11$), the control group, was given no inspection at any time, but spent 1 min. between the pretest and the posttest standing at rest between the inspection bars without touching them. Group 2 ($N = 14$), the 1-min. group, was given 1 min. of inspection between the pretest and the posttest. Group 3 ($N = 14$), the 5-min. group, was given 1 min. of inspection between the pretest and the posttest, and 4 min. of further inspection following the posttest. All *S*s were 1956 summer school students at the University of Colorado and were unpaid volunteers. Groups 1 and 2 were run concurrently, with alternate assignment of *S*s

[1] This study was supported in part by a grant from the Council on Research and Creative Work of the University of Colorado.

[2] Now at the School of Medicine and Dentistry, University of Rochester.

and the groups equally weighted for sex. Each group contained about twice as many men as women; the age range was from 18 to 56. Group 1 initially had an N of 14, but 3 *Ss* had to be dropped because they failed to appear on one of the five days.

If amount of residual FAE is a function of amount of inspection, Group 3 (5 min. inspection daily) should show the greatest increase in pre-inspection measures from day to day, Group 2 (1 min. inspection daily) should show a lesser increase, and Group 1 (no inspection) should show no increase.

Results and Discussion

As Fig. 1 shows, the predictions seem to hold. There is an increase in pre-inspection match (indicating more and more "permanent satiation," or an ever larger residual FAE relative to the pre-inspection trials on Day 1) as a function of day, with Group 3, the 5-min. group, exhibiting the greatest increase and Group 1, the control group, the smallest increase.

Although the family of curves on the kinesthetic FAE seems well ordered, with each curve in its expected place relative to the other two, the increase ($P = .02$, by t test for correlated groups) in the matches made by Group 1 (the control group) looks, at first sight, inexplicable. Since there is no inspection, there should be no increase and the curve should be a horizontal line. However, an examination of details of the procedure suggests a possible explanation. The test matches, both pre-inspection and post-inspection, were made with two trials (Trials 1 and 4) starting at a width slightly below that of the standard, while the other two trials (Trials 2 and 3) were started at a width considerably greater than that of the standard. Therefore during the width matches, the right hand was exposed more to widths greater than the standard than to widths less than the standard. In other words, the right hand received some "inspection" of widths greater than the standard at the time that test matches were made. It is quite possible that this inspection during test matches was enough to induce a small FAE, and hence produced the small residual FAE evidenced by Group 1's increase in variable width matched to the standard width. The results would then be interpreted as follows: With a relatively small amount of daily inspection, a small residual FAE is generated (Group 1, control). With 1 min. of daily inspection superimposed upon this slight inspection also experienced by the control group, the residual FAE grows more rapidly (Group 2, 1 min.). With 5 min. of daily inspection in addition to the amount induced by the measuring technique, the residual FAE grows even more strongly (Group 3, 5 min.).

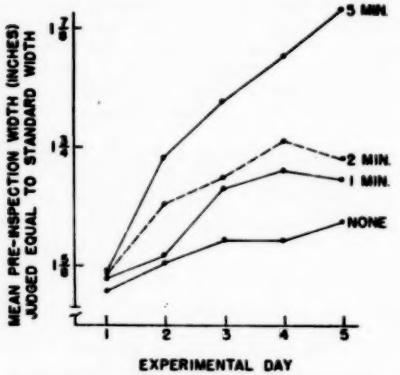

Fig. 1. Growth of residual FAE over time with varying amounts of daily inspection. Parameter: amount of daily inspection. The "2 min." curve is from an earlier exploratory study.

An earlier study (6), performed for other purposes, yielded data that could be reanalyzed in the light of the present study. Essentially 9 *Ss* were given 2 min. of inspection daily for several successive days, using the same kines-

thetic FAE apparatus as that in the present study, although the technique was slightly different. Plotting the curve of mean pre-inspection match as a function of day for this 2 min. group yielded a function which fit well into the family of curves in Fig. 1. For each of five successive experimental days, the mean for the 2-min. group fell between the 5-min. and the 1-min. curves obtained in the present study (see the dotted curve in Fig. 1).

Some further measures, not previously reported, of the longevity of residual FAE were performed on several *S*s in the earlier study (**6**), by obtaining post-inspection matches as long as 6 mo. after the last inspection. Seven of the nine that were available for repeat measures showed pre-inspection matches 8 wk. after the last inspection significantly greater—at the .05 level, by *t* test for each individual *S*—than the pre-inspection match on Day 1. All of the 6 *S*s on whom it was possible to obtain measures 16 wk. after last inspection gave matches, without further inspection, that were still greater than their pre-inspection matches on the first day. Of 5 *S*s on whom it was possible to obtain matches 24 wk. after last inspection, 4 still had matches which were higher than the pre-inspection measures on Day 1. Finally, both of 2 *S*s on whom measures were available 32 wk. after last inspection still were significantly above the pre-inspection matches on Day 1. Apparently the residual FAE can be measured as long as ¾ yr. after final inspection. It must be remembered, however, that the bimonthly tests may in part have been responsible for the continued elevation of the matches; but even if the measuring technique itself produced some appreciable inspection, the residual effect of this inspection could be measured at least 2 mo. later.

Summary

Kinesthetic FAE measures were obtained on three groups for five successive days, with the groups varying in the amount of daily inspection. Evidence for significant residual FAE was obtained by plotting daily pre-inspection matches. The greater the daily inspection, the greater the residual FAE 24 hr. later; the rate of growth of the residual FAE also varied monotonically with the amount of daily inspection. Data from another study suggested that the residual FAE may be measurable even ¾ yr. after last inspection. The data can be interpreted to demonstrate that satiation effects may persist lawfully over relatively long periods of time.

References

1. Hammer, E. R. Temporal factors in figural after-effects. *Amer. J. Psychol.*, 1949, **62**, 337–354.
2. Köhler, W., & Dinnerstein, D. Figural after-effects in kinesthesis. In *Miscellanea psychologica Albert Michotte*. Louvain: Institut Superieur de Philosophie, 1947, 196–220.
3. Köhler, W., & Fishback, J. The destruction of the Müller-Lyer illusion in repeated trials. I and II. *J. exp. Psychol.*, 1950, **40**, 267–281, 398–410.
4. Köhler, W., & Wallach, H. Figural after-effects: an investigation of visual processes. *Proc. Amer. Philos. Soc.*, 1944, **88**, 269–357.
5. Wertheimer, M. Constant errors in the measurement of figural after-effects. *Amer. J. Psychol.*, 1954, **67**, 543–546.
6. Wertheimer, M., Levine, H., & Wertheimer, N. The effect of experimentally induced changes in metabolism on perceptual measures of metabolic efficiency. *Percept. mot. Skills*, 1955, **5**, 173–176.

(Received March 26, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

TRANSFER TO A MOTOR TASK AS INFLUENCED BY CONDITIONS AND DEGREE OF PRIOR DISCRIMINATION TRAINING[1]

ALBERT E. GOSS AND NORMAN GREENFELD[2]

University of Massachusetts

Various conditions of verbal pretraining and stimulus predifferentiation apparently facilitate subsequent discriminative behavior (**1, 2, 3, 4, 7, 10, 12, 14, 16, 17**).[3] Several explanatory mechanisms for such facilitation have been proposed. In an attempt to ascertain the relative adequacy of certain of the proposed explanatory mechanisms, this study investigated acquisition of discriminative motor responses as a function of degree of mastery of seven different conditions of prior discrimination training.

Explanatory Mechanisms

Acquired distinctiveness of cues.—One explanation of instances of positive transfer is that prior discrimination training assures the occurrence of distinctive verbal response-produced stimuli, whose presence increases the dissimilarity of the stimulus compounds to be discriminated (**5, 8**). In several studies in which verbal pretraining and stimulus predifferentiation experiences have improved subsequent discrimination behavior, determination of the contribution of this acquired distinctiveness of cues has been hampered by lack of controls for receptor-orienting responses, learning to pay attention to stimuli, or "learning set" (**1, 3, 9**). The results of other studies in which those factors were controlled have been inconclusive as to whether experimental strengthening of discriminative verbal responses occasioned positive transfer over and above that brought about by activation of pre-experimentally acquired discriminative verbal responses (**7, 10, 16**). But greater achievement due to the latter source would also be consistent with the mechanism of acquired distinctiveness.

Some investigators, however, have failed to obtain positive transfer with one or more different combinations of conditions of verbal discrimination training and of subsequent discriminative verbal or motor responses (**1, 2, 9, 12, 14, 17**). Several reasons for both inconclusive and negative results now seem plausible. First, the discriminative verbal responses may not have always been strengthened sufficiently (**1, 2, 14**). Second, dissimilarity of the prior training and the test situations may have reduced the strengths of the discriminative verbal responses in the latter task (**14**). Further, when the sets of discriminative responses of both verbal pretraining and transfer tasks have the same general topography (e.g., if both require selection and then saying letters of the alphabet [**9**]), care must be taken to avoid negative transfer due to conflict between the discriminative responses of the first task and those to be acquired in the subsequent test for transfer.

During pretraining experiences to control for exposure to stimuli of the transfer task, already-established repertoires of discriminative responses to those stimuli may have been activated and possibly strengthened. Or such responses may

[1] This study was supported by a grant from the American Philosophical Society. Robert Tanofsky aided in the collection of data, and Betsy Silun and Sally Grahn contributed to the statistical analyses.

[2] Now at the New York State College for Teachers in Albany.

[3] These and many subsequent references are illustrative. Additional relevant studies are cited in **2, 7, 12, 17**.

have been aroused during the first few trials of the transfer task. In either case activation of already-acquired discriminative verbal responses to stimuli of the transfer task may have been sufficient to offset any advantages stemming from experimental strengthening of discriminative verbal responses. Other factors which may account for failure to observe facilitation are (*a*) the set to respond rapidly (**14**), (*b*) marked complexity of the motor task (**2**), (*c*) use of tasks which emphasize "motor" rather than perceptual processes (**12**), (*d*) a criterion task requiring discriminative judgments (**1**), and (*e*) if the discrimination task is too easy, enhanced distinctiveness resulting from dissimilar response-produced stimuli might be a superfluous aid (**8**).

Other explanatory mechanisms.—In addition to the acquired distinctiveness of cues it is possible that strengthening of receptor-orienting responses to the most discriminable aspects of stimuli during discrimination training may have been an alternative or supplementary reason for positive transfer to the test tasks (**11**). The *S*s may then give different names or identifications to the more discriminable parts (**8, 15, 16**). Furthermore, warm-up—in the form of learning to attend to stimuli or "learning to learn"—may take place during verbal pretraining, thus bringing about positive transfer to later tasks (**1, 2, 3, 4, 7, 10, 14, 16, 17**). And, as noted previously, many *S*s may have learned discriminative names for task stimuli prior to their experimental experiences which are then activated during discrimination training to serve as the actual source of any subsequent facilitation (**7, 10**). One problem which this proposed source of facilitation introduces is that the nature and course of changes in such responses during discrimination training have usually been inferred rather than directly observed. In addition, facilitation attributable to experimental strengthening of discriminative responses should then be measured against a baseline of positive transfer due to activation of already available responses.

Also accorded explanatory significance have been (*a*) knowledge of relevant stimulus attributes (**1**); (*b*) "learning . . . to use these stimuli as cues for action" (**18**, p. 759); (*c*) avoidance of "confusing stimuli perceptually" (**9**, p. 161); (*d*) a "general labeling *process*" (**9**, p. 166); and (*e*) "a context which provided an objective" (**9**, p. 167). Before the influence of these factors can be assessed, however, further specification in terms of stimulus-response principles appears to be required.

Rationale for the Experimental Design

To meet various of the shortcomings of previous investigations, in the present study external stimuli appeared in one place; they did not involve form; they differed along only one dimension (intensity of white light); and they were sufficiently similar that differential reinforcement was necessary to bring verbal discrimination to a high level. Also, it was presumed that the criterion task used here emphasized "perceptual" processes, that it was not inherently difficult, and that the to-be-learned manipulative responses were compatible with the discriminative verbal responses. Controls were introduced for the possible effects of activation of discriminative names brought by *S*s to the experiment, of "learning set," and of attending to the stimuli.

Seven conditions of prior discrimination training were used in this study. Three of the seven conditions involved the strengthening of discriminative verbal responses by means of experimentally controlled differential reinforcement. Nonsense syllables were the responses in one condition; the other two conditions involved familiar names supplied by *E* and familiar names selected from each *S*'s repertoire. These latter two conditions were introduced to see whether

sets of responses which differed somewhat from each other and from nonsense syllables with respect to source, form, and initial strengths would influence the amount of positive transfer to the criterion task. Any attempt to separate the confounded variables of source, form, and initial strengths, however, seemed premature. In the condition in which each *S* supplied familiar names, reinforcement by *E* saying "right" was substituted for the paired-associates procedure employed with the other two conditions.

Three additional conditions of the study were aimed at providing decreasing levels of activation of discriminative names already in *S*s' repertoires. Instructions for the first of these were for *S*s to look at or "see" the stimuli, discriminate among them, and learn to give them different names overtly. Overt occurrence of pre-experimentally acquired responses permitted assessment of possible upward trends in the accuracy of such responses. Also, any effects of saying pre-experimentally acquired labels aloud could be determined by comparison with the second condition, wherein *S*s were given the same instructions except that verbalization was to be covert. In addition to activation of previously acquired verbal discriminations, both conditions presumably involved a "learning set" without, however, accompanying *E*-manipulated differential reinforcement. The third of these conditions, which involved instructions to see and discriminate among the stimuli, was viewed as establishing a condition of even less activation of already available naming responses.

A final condition allowed for experience in attending to the stimuli with, presumably, minimum arousal of discriminative labels. Serving as the baseline for these seven conditions of prior discrimination training was a control group which learned the manipulative responses without prior experimental exposure to the task stimuli.

Seven degrees of mastery were employed in order to obtain a description of the relationship between extent of facilitation and amount of prior training under each of the conditions except the one involving nonsense syllables for which there were only six degrees. Thus there were 48 different experimental combinations of conditions and degrees and a control group (no prior training).

Method

Subjects.—There were 10 *S*s in each of the 48 experimental groups and 20 in the control group. All were students in the introductory psychology course at the University of Massachusetts. Their assignment to the groups was random within the limitation that each group consist of approximately half men and half women.

Apparatus and stimulus materials.—The verbal-motor discrimination apparatus consisted of two circular apertures in a vertical panel for presentation of light and word stimuli, and a 4-in.-high lever for manipulative responses placed vertically in the center of a horizontal panel. The light stimuli were presented in the lower of the two apertures and the stimuli for verbal responses in the upper one. The lever, which replaced the four toggle switches used earlier (**7**), was pivoted for movement downward in slots extending forward, backward, left, and right. Completion of the downward movement resulted in pressure against a button at the end of each slot. Depression of each button activated a microswitch which lit bulbs on the back of the apparatus to indicate *S*'s choice to *E*. If the choice was correct, the stimulus light was automatically turned off.

In order to increase the difficulty of the discrimination, as well as to reduce the glare of the most intense light, intensity of the light at the position of *S*'s eyes was reduced from the 3.40 ft.-candles of previous studies (**7, 10**) to 1.80 ft.-candles. Average readings for the other intensities of white light were approximately 1.50, 1.30, and 1.10 ft.-candles.

TABLE 1

SUMMARY OF CONDITIONS AND DEGREES OF PRIOR TRAINING

Prior Training Conditions	Code	Degrees (Criteria of mastery or numbers of trials)						
Learn different nonsense syllable responses to similar intensities	NS	3/4	6/8	9/12	11/12	100%	200%	—
Learn different familiar word responses supplied by *E*	FWES	3/4	6/8	9/12	11/12	100%	200%	M*
Learn different familiar word responses supplied by each *S*	FWSS	3/4	6/8	9/12	11/12	100%	200%	M
Instructed to look at, discriminate among, and give different names to stimuli overtly	SDNO	32	44	56	68	108	132	M
Instructed to look at, discriminate among, and give different names to stimuli covertly	SDNC	32	44	56	68	108	132	M
Instructed to look at and discriminate among the stimuli	SD	32	44	56	68	108	132	M
Instructed to see or look at the stimuli	S	32	44	56	68	108	132	M
Control	C	0						

* 172 trials to match those given to the NS-200% group.

Procedures for discrimination training.—Table 1 summarizes the conditions of prior training and the amount of training for each of the 49 groups. The three verbal-learning conditions involved, respectively, differential strengthening of nonsense syllable labels (NS), or of familiar words supplied by *E* (FWES), or by each *S* (FWSS). The nonsense syllables *jer*, *vol*, *cuf*, and *nig*, were paired with decreasing intensities in that order. *Very bright*, *bright*, *dull*, and *very dull* were the familiar word labels for decreasing intensities which were supplied by *E*.

For the three verbal-learning conditions, and also for the other four conditions, the control group defined the lowest or zero level of mastery. The six additional criteria of increasing mastery for each verbal-learning condition were 3 correct responses in a 4-trial block (3/4), 6 correct responses in two successive 4-trial blocks (6/8), 9 correct responses in three successive 4-trial blocks (9/12), 11 correct responses in three successive 4-trial blocks (11/12), 100% overlearning based on two times the number of trials to the 11/12 criterion (100%), and 200% overlearning based on three times the number of trials to the 11/12 criterion (200%). However, in order to avoid differential elimination of slow learners from the 11/12, 100%, and 200% criterion groups, alternative criteria of 96, 144, and 192 trials, respectively, were used for some *S*s in those groups. Because markedly more trials were required to learn nonsense syllables than familiar words to the 200% criterion, a matching condition (M) was introduced to equate all conditions with the NS-200% group in terms of number of trials. To do so 172 trials were administered to an additional group within each of the other six training conditions.

Learning in the NS and FWES conditions was by the paired-associates technique with lights presented for 2 sec. followed by light-syllable or light-familiar word pairs for 2 sec. Interpair intervals were 4 sec.

The FWSS condition was obtained by instructing *S*s to look at the stimuli and try to give each different light a different name which, if it were correct, *E* would designate as "right." After *E* had indicated which was the "right" name for a given light, they were to continue to give that name to that stimulus. Because *S*s had to see the stimuli and try out some names in order to produce sufficient different responses for *E* to select four for differential reinforcement, no responses were reinforced during the first 12 trials. Labels given to each intensity during those trials were recorded, and, on Trial 13, *E* began to select from among the labels and differentially reinforce four of them by saying "right" for only one name for each intensity. Whenever possible, *E* selected the particular names for each intensity so that they would conform to familiar ways of describing decreasing intensities (e.g., dim, medium, medium bright, bright), or of ordering events (e.g., A, B, C, D). As before (**10**), however, some *S*s gave sets of labels, such as personal names, which could not be ordered in this fashion. The duration of each intensity on each trial for this condition, and also for the conditions described below, was 4 sec., with 4-sec. interstimulus intervals.

In Cond. SDNO (Seeing, Discriminating, Naming Overtly), Ss were told to look at the lights, discriminate among them, and to provide different names for them which were to be said aloud. These verbalizations were recorded. Instructions for Cond. SDNC (Seeing, Discriminating, Naming Covertly) were the same except for the stipulation that the names need not be said aloud. The third activation condition (Cond. SD; Seeing, Discriminating) required Ss to look at and discriminate among the lights but discriminative naming was not mentioned. Upon completion of SDNO, SDNC, or SD experiences Ss were asked how the lights they had seen differed and how many different ones had occurred. Under Cond. S (Seeing) discrimination instructions were omitted altogether in order that Ss would have experience in attending to the stimuli with minimum arousals of a "learning set" or a set to develop or elicit names.

The number of trials administered to Ss in these conditions were intended to equal the pooled means of trials required by all three verbal-learning conditions to attain each of the successively more stringent criteria of mastery. In practice they somewhat exceeded the values they were to match. Specific trial values were 32 to match 3/4, 44 for 6/8, 56 for 9/12, 68 for 11/12, 108 for 100%, and 132 for 200%. The 172 trials of the *M* groups matched those for the NS-200% group.

Procedure for manipulative task.—Approximately 1 min. was required to introduce Ss to the criterion task which was to push the lever in a different direction for each intensity. Correct responses—which, Ss were informed, turned off the light—were movement of the lever to the right for the very dull light, backward for the bright light, left for the very bright light, and forward for the dull light. The same instructions were given to the controls who had not previously seen the lights. Intertrial intervals were not held constant; they averaged about 4 sec.

Two changes from procedures of the most closely related previous studies (**7, 10**) were made. A noncorrection method was used, instead of a correction method, and the number of trials was increased from 48 to 72.

Results

Discrimination training.—Means and *SD*'s of the total number of trials for each level of mastery of responses acquired under Cond. NS, FWES, and FWSS are summarized in Table 2. With some exceptions, attainment of the successively more stringent criteria required increasing numbers of trials, and familiar words, particularly those supplied by *E*, were learned more rapidly than nonsense syllables.

Condition SDNO was included in order to determine the extent to which Ss would be able to use preexperimentally acquired verbal responses to discriminate among the four intensities. Also of interest was whether such discriminations would improve with increasing amounts of

TABLE 2

Means and *SD*'s of Trials, Including Criterial Trials
(*N* = 10 for each group)

Criteria of Mastery	Conditions					
	NS		FWES		FWSS*	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
3/4	28.4	9.9	13.2	3.6	36.0	24.7
6/8	51.2	17.2	20.0	8.2	41.2	23.7
9/12	55.2	23.7	26.4	8.0	50.0	23.3
11/12	62.0	21.3	54.0	23.9	82.0	22.6
100%	108.8	49.9	86.0	36.3	134.4	49.7
200%	156.0	42.2	81.6	37.1	95.2	51.1
M	—	—	172.0	—	172.0	—

* Means include the first 12 trials during which there was no differential reinforcement of responses.

TABLE 3

MEANS AND *SD*'S OF CORRECT RESPONSES FOR THE FIRST 24 AND ALL 72 MOTOR LEARNING TRIALS

Cond.	Degrees (Criteria or trials)													
	3/4 or 32		6/8 or 44		9/12 or 56		11/12 or 68		100% or 108		200% or 132		*M*	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
First 24 Trials														
NS	9.5	4.8	8.0	2.1	9.6	1.8	11.3*	3.4	9.0	4.1	11.4*	5.3	—	—
FWES	7.6	3.5	9.1	4.7	8.3	4.6	9.2	2.7	10.8*	4.8	11.1*	4.7	8.2	2.4
FWSS	10.0*	2.5	7.6	2.8	8.2	4.3	9.9	3.7	10.5*	4.4	9.9	4.2	9.7	3.8
SDNO	8.4	3.6	8.7	2.1	9.9	2.9	6.2	3.0	7.9	3.2	9.2	4.0	7.0	2.8
SDNC	10.3*	5.8	7.4	3.9	6.9	3.6	8.4	2.6	7.3	4.0	7.1	3.7	9.2	2.7
SD	8.0	2.2	7.7	3.6	9.2	3.4	10.6*	3.8	7.4	3.0	10.9*	3.7	10.2*	4.1
S	8.0	3.7	6.5	2.2	6.5	3.0	7.8	2.2	7.6	3.2	8.8	2.7	8.7	3.1
All 72 Trials														
NS	31.3	9.0	34.7*	5.0	37.0*	6.4	46.0*	10.3	39.4*	13.1	41.6*	12.0	—	—
FWES	34.5*	7.7	32.6*	11.7	31.3	12.5	37.8*	8.0	42.5*	9.6	44.2*	11.5	37.1*	6.4
FWSS	43.0*	5.7	31.2	13.2	35.2*	13.2	38.6*	11.8	42.9*	13.0	37.5*	13.9	42.6*	11.0
SDNO	30.1	9.1	34.1*	5.9	34.3*	10.3	28.8	8.8	30.6	9.6	36.8*	9.9	27.9	7.6
SDNC	37.4*	13.5	29.4	10.9	29.9	13.0	33.9*	10.7	31.4	12.4	30.6	10.8	36.2*	10.9
SD	32.0	8.8	34.3*	13.2	33.3*	9.6	36.2*	12.6	31.4	9.0	34.9*	12.0	37.9*	8.0
S	28.6	10.1	26.4	12.9	22.1	11.4	28.6	7.2	26.0	14.3	30.5	12.3	30.0	12.1

* Different from Cond. C means at .01 level for first 24 and for all 72 trials for 457 *df*.

experience with the intensities. The overt naming responses of the seven SDNO groups were therefore recorded and analyzed to see how many *Ss* had reached a criterion of responding to each intensity with a distinctive name on two or three of its three occurrences during the last 12 trials for each amount of experience. Such discriminative responses were made by 6 of the 70 *Ss* in Cond. SDNO; and 3, 2, 1, and 0 lights were so labeled with distinctive names by 24, 32, 6, and 2 *Ss*, respectively. Thus *Ss* instructed to respond differentially to the intensities with names already in their repertoires were able to do so, but at an average level which was lower than the 9/12 criterion. Although the *N*'s of 10 for the separate groups were too small for reliable conclusions, increased experience did not seem to produce an upward trend in the levels of accuracy of these activated repertoires of discriminative responses.

Acquisition of manipulative responses.—Table 3 presents means and *SD*'s of correct responses for the first 24 trials and for all 72 trials of learning differential manipulative responses. Means for Cond. C for the first 24 and for all 72 trials were 7.3 ($SD = 1.74$) and 26.4 ($SD = 5.65$), respectively. All of the means of the verbal learning conditions (NS, FWES, FWSS), both for the first 24 and for all 72 trials, were above the corresponding means of Cond. C. For the first 24 trials means of two of the SNDO groups (68, M), three of the SDNC groups (56, 108, 132), and none of the SD groups were less than or equal to the

mean of Cond. C. All of the 21 72-trial means of the SDNO, SDNC, and SD conditions were larger than the C mean of 26.4. Two of the means of the S groups for the first 24 trials and three of those for all 72 trials were smaller than the corresponding means of the controls. The over-all pattern of these differences thus suggested (*a*) an initial and continuing superiority of the verbal-learning conditions to Cond. C, (*b*) less marked differences initially, but for all 72 trials more pronounced superiority of the seeing and discriminating conditions (SDNO, SDNC, SD) to the controls, and (*c*) approximate equality of Cond. S and C both for the first 24 and for all 72 trials.

The *F* of 2.80 for Conditions of the Conditions × Degrees analysis of variance (Table 4) for just the first 24 trials was significant at the .05 level. Further comparisons of pairs of means for all degrees of each of the conditions by *t*'s, and a nonsignificant *F* for Conditions when the S groups were removed, indicated that the effect of this variable was due entirely to the inferior performance of the S groups. Therefore it was concluded that during the first 24 trials, the average levels of performance of all training conditions but S were essentially the same.

The error term of the analysis of variance ($N = 10$; $df = 438$) and the *SD* of Cond. C were used to obtain the difference between the mean of 7.3 for Cond. C and Trial 1–24 means of each of 48 training groups necessary for significance at the .01 level (**13**). This value was 2.68. Two each of the NS (11/12, 200%), FWES (100%, 200%), and FWSS (3/4, 100%) groups, one of the SDNC (32) groups, three of the SD (68, 132, M) groups, and none of the SDNO and S groups exceeded the Cond. C mean by this amount. The significantly superior values are the Trial 1–24 means in Table 3 with asterisks.

TABLE 4

ANALYSIS OF VARIANCE FOR CONDITIONS AND DEGREES FOR THE FIRST 24 TRIALS

Source	*df*	*MS*	*F*
Conditions (C)	6	39.96	2.80*
Degrees (D)	5	24.90	1.75
C × D	30	11.06	—
Error	438	14.26	

* $P < .05$.

TABLE 5

ANALYSIS OF VARIANCE FOR CONDITIONS, DEGREES, AND 24-TRIAL BLOCKS

Source	*df*	*MS*	*F*
Conditions (C)	6	377.93	9.02*
Degrees (D)	5	83.04	1.98
C × D	30	42.44	1.01
Error (*b*)	438	41.91	
Blocks (B)	2	2719.08	358.72*
C × B	12	26.85	3.54*
D × B	10	9.33	1.23
C × D × B	60	6.86	—
Error (*w*)	876	7.58	

* $P < .01$.

The analysis of variance for means based on all 72 trials is summarized in Table 5. Because *t*'s had indicated that means of the 200% and M criteria within any of the conditions did not differ significantly, scores for *S*s in these two criteria groups for each training condition were pooled so that comparisons were made for six rather than seven degrees of experience under the seven conditions. Blocks of trials and its interactions with the other variables were the "within" sources of variance (**13**).

Conditions, Blocks, and Conditions × Blocks yielded *F*'s significant at beyond the .01 level. The more specific differences contributing to the

significant *F* for Conditions were investigated by *t*'s whose error variances were based on the "error (*b*)" mean square. These analyses indicated that the three verbal-learning conditions, which did not differ from each other, were superior to the SDNO, SDNC, and SD conditions, which also did not differ from each other. In turn, the seeing and discriminating conditions resulted in more correct responses than Cond. S.

The difference between the Cond. C mean and the Trial 1–72 means of the training groups required for significance at the .01 level was based on a *t* whose error variance was computed from the "error (*b*)" mean square ($N = 10$; $df = 438$) from Table 5 and the Cond. C *SD* of 5.65 for all 72 trials (**13**). This value was 6.14. Those means for all 72 trials in Table 3 which exceeded the Cond. C mean of 26.4 by 6.14 are designated by asterisks. Differences of less than this amount were obtained for one each of the NS (3/4), FWES (9/12), and FWSS (6/8) groups, for four each of the SDNO (32, 68, 108, M) and SDNC (44, 56, 108, 132) groups, for two of the SD (32, 108) groups, and for all seven of the S groups. Means of the separate groups of each of the discrimination training conditions were summed and averaged for all degrees of mastery or experience and then compared with the mean of Cond. C. The differences required for significance at the .01 level were again computed from the "error (*b*)" mean square (*N*'s = 70 or 60) and the *SD* for Cond. C. All of the differences except that between Cond. S and Cond. C exceeded those of 4.44 for Cond. C with FWES, FWSS, SDNO, SDNC, SD, and S, and of 4.54 for Cond. C and NS. Thus there is little doubt that all of the discrimination training conditions but seeing the intensities led to significant facilitation of motor learning.

The significant *F* for Blocks merely indicated that learning occurred during the 72 motor discrimination trials. As noted above, all of the training conditions but S were at the same level during Trials 1–24. During the next two 24-trial blocks the NS, FWES, and FWSS conditions acquired the manipulative responses more rapidly than the SDNO, SDNC, and SD conditions which were learning faster than the S groups. The significant *F* for Conditions × Blocks reflects these differences in rates of strengthening of the manipulative responses.

An additional analysis of variance (Table 6) indicated that neither the main effects of Cond. NS, FWES, and FWSS nor their interaction with degrees was significant. Means for each degree of mastery of these conditions were therefore pooled and plotted as the upper curve of Fig. 1. Similarly the *F*'s for the main effects of Cond. SDNO, SDNC, and SD and the Conditions × Degrees interaction were not significant (Table 6). The means of these conditions for each amount of experience were averaged to obtain the middle curve. Because

TABLE 6

ANALYSES OF VARIANCE FOR CORRECT RESPONSES FOR ALL 72 TRIALS FOR COND. FWES, FWSS, AND NS AND FOR COND. SDNO, SDNC, AND SD

Source	FWES, FWSS, and NS			SDNO, SDNC, and SD		
	df	*MS*	*F*	*df*	*MS*	*F*
Cond. (C)	2	46.64	—	2	116.62	1.15
Deg. (D)	5	443.99	3.54*	5	5.03	—
C × D	10	136.82	1.08	10	96.67	—
Error	182	125.52		192	101.70	

* $P < .01$.

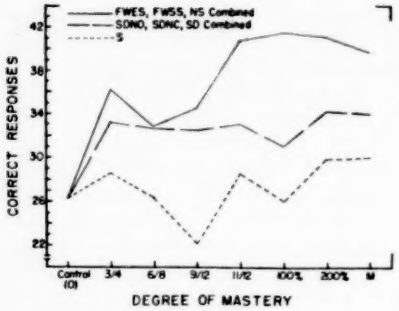

FIG. 1. Mean correct responses for increasing degrees of mastery or amounts of pooled FWES, FWSS, NS conditions, pooled SDNO, SDNC, SD conditions, and Cond. S.

means for each amount of Cond. S fell below those for the other conditions they have been plotted separately.

Inspection of Table 3 and Fig. 1 suggests that degree of experience affected acquisition of manipulative responses in only a small part—primarily between the 9/12 and 11/12 points for both separate and pooled NS, FWES, FWSS conditions—of the $6 \times 7 \times 3$ matrix of the analysis of variance of Table 5. Therefore, that neither the main effect of this variable nor its interaction with verbal training conditions was significant is not surprising. Because results for the verbal learning conditions were consistent with theoretical expectations (**8**), however, the effects of degree of verbal learning only (Cond. NS, FWES, FWSS) were tested (Table 6). The significant F suggests that amount of facilitation was directly related to degree of mastery of the experimentally strengthened discriminative verbal responses. Further comparisons with t's indicated that the only significant differences between pairs of means for each degree of the three conditions combined were the comparisons of the 3/4, 6/8, and 9/12 criteria with each of the more stringent criteria. The F's for different amounts of SDNO, SDNC, and SD experiences (Table 6) and, although not shown here, of Cond. S, were not significant.

DISCUSSION

Discrimination Training

The discriminative verbal responses to the light intensities which were learned to increasing degrees of mastery were nonsense syllables, familiar words supplied by *E*, or labels selected from *S*s' repertoires. In general, as reported previously (**10**), nonsense syllable responses were learned to each criterion at slower rates than the familiar word discriminations.

Discrimination among the intensities resulting from varying amounts of experience in using pre-experimentally acquired verbal responses was investigated by means of a condition in which *S*s were instructed to look at the lights, discriminate among them, and give each a different name overtly. The findings for this condition confirmed earlier observations (**10**) that *S*s so instructed were able to respond to the lights differentially with names acquired prior to their experimental exposure to the intensities. The levels of discriminative accuracy attained, however, were lower than a 9/12 learning criterion, and there was no evidence that increasing amounts of experience in trying to respond to the different lights with different names brought about greater accuracy. Without further differential reinforcement, therefore, activation of pre-experimentally acquired discriminative verbal responses was apparently complete by 32 trials and, for relatively similar light stimuli, fell short of higher levels of discriminative proficiency.

Acquisition of Motor Responses

Conditions.—At least for 11/12 and greater degrees of mastery, acquisition of nonsense syllable names and of

familiar word labels (whether supplied by *E* or by *S*) yielded significantly greater transfer to the motor discrimination than seeing, discriminating, and naming overtly or covertly, or of seeing and discriminating without explicit instructions to name discriminatively. The three conditions involving seeing and discriminating also produced positive transfer in contrast to seeing the stimuli, which had no facilitative influence.

The obtained superiority of the three verbal learning conditions was more striking than that reported by Goss (**7**) in an earlier study, and it contrasted with the nonsignificant differences among five discrimination training conditions reported by Holton and Goss (**10**). The more difficult discrimination required plus, perhaps, the noncorrection procedure seem the most likely sources of differences between the present and earlier observations with similar apparatus. In general, differences between the present findings and results of other studies, particularly those in which prior discrimination training produced no facilitation, probably reflect one or more of the factors noted in the introductory section, most of which were eliminated or minimized by the apparatus and procedures of this experiment.

There were no differences in motor learning associated with the three conditions of verbal learning. Possibly so long as verbal response-produced stimuli are dissimilar, amount of facilitation will be equal. Alternatively, the observed homogeneity may be due to use of particular but not yet completely specifiable combinations of sources, forms, initial strengths, and modes of reinforcement of the three sets of responses. The seeing, discriminating, naming-overtly condition and, as will be noted below, probably the seeing, discriminating, naming-overtly and seeing, discriminating conditions involved activation of already acquired discriminative responses. Such activation was necessarily confounded with warm-up in the form of "learning set" and experience in attending to the stimuli. The seeing condition controlled for attending to the stimuli. Therefore, it seems warranted to conclude that experimental strengthening of discriminative verbal responses had facilitative effects over and above positive transfer attributable to activation and warm-up. The apparatus and stimuli eliminated the factors of relevant-irrelevant attributes of stimuli and enhanced distinctiveness due to receptor-orienting responses.

In the seeing, discriminating, naming-overtly condition the intensities were named with some differential consistency. On the basis of comparable results for learning the manipulative task the occurrence of similar consistent verbal responses, although they were not observed, can be postulated for the conditions that involved seeing, discriminating, and covert naming and seeing and discriminating with no instructions to name. Therefore, the more rapid acquisition of the motor task associated with these conditions than with seeing the stimuli lends some support to the hypothesis of acquired distinctiveness based on dissimilar response-produced stimuli. But "learning set" may have been a contributing or crucial factor.

Conditions of seeing, discriminating among, and naming the intensities overtly or covertly were no more effective than the condition in which the instructions were to see and discriminate. The three types of experience evidently did not involve different levels of activation of pre-experimentally acquired discriminative labels. Little is known about the consequences of overt vs. covert vocalization of well-learned responses, but because of extensive experience with both levels of response amplitude it would not be surprising that the effect is negligible. Also, instructions to name stimuli discriminatively and to discriminate among them may be redundant, since any discriminating by adult humans almost certainly involves some sort of discriminative verbal responses. Therefore the possible difference between the two conditions in which naming was stimulated and Cond. SD would lie in the degree

and accuracy of *Ss*' use of pre-experimentally acquired names. Again, with well-learned responses, there might be no difference.

Degree.—Primarily because of differences between the three lower and the 11/12 and more stringent criteria, degree of mastery of familiar word or nonsense syllable responses influenced rate of motor learning. Due to the more difficult discrimination required and/or the noncorrection procedure presumably, the increase from the 9/12 to 11/12 points was more marked than that obtained by Goss (**7**).

While differences among Holton and Goss' (**10**) groups and those of this study who had learned to, or exceeded, the 11/12 criterion were not significant, Cantor (**4**) found some evidence of a decline in positive transfer between four and 12 blocks of 32 discrimination training trials each.

The means for the motor learning of the verbal-learning conditions to the 3/4 criterion were higher than the performance of the controls. Therefore, even lower levels of mastery are necessary to determine the form of the curve declining to the level of the control or no experience level. Regardless of the "distance" between the 0 and 3/4 points along the abscissa, such curves would be approximately sigmoidal in form (**8**).

As noted above, analysis of the overt responses of the seeing, discriminating, naming-overtly groups indicated that without differential reinforcement maximum discriminative accuracy is reached in 32 trials. It is not surprising, therefore, that amount of seeing and discriminating experiences was not related to extent of positive transfer.

Summary

The experiment was concerned with transfer to a discriminative motor task as a function of degree of experience with different conditions of prior discrimination training with stimuli of the manipulative task. There were 48 experimental groups of 10 undergraduates each and a control group of 20. Four white lights of different intensities served as stimuli to which different familiar words supplied by *E*, *S*-supplied familiar words, or nonsense syllables were learned to six or seven criteria of increasing mastery. Other *Ss* were given equated amounts of experience in seeing, discriminating, and naming the stimuli either overtly or covertly, seeing and discriminating, or seeing. All *Ss* then had 72 trials to learn to move a lever in a different direction for each intensity.

With exception of the seeing experience, all types of activation or acquisition of discriminative verbal responses resulted in some positive transfer. There were no differences among the three types of verbal-learning experience which, for higher degrees of mastery, were superior to the three types of experiences which involved seeing and discriminating. Amount of positive transfer increased with mastery of verbal responses but did not vary with amount of any of the seeing and discriminating experiences. The over-all pattern of the results was considered consistent with the hypothesis that dissimilar verbal response-produced stimuli increased the distinctiveness of intensities.

References

1. Arnoult, M. D. The transfer of predifferentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45,** 401–409.
2. Battig, W. F. Transfer from verbal training to motor performance as a function of motor task complexity. *J. exp. Psychol.*, 1956, **51,** 371–378.
3. Cantor, G. M. Effects of three types of pretraining on discrimination learning in preschool children. *J. exp. Psychol.*, 1955, **49,** 339–342.
4. Cantor, J. H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *J. exp. Psychol.*, 1955, **50,** 180–184.
5. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
6. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47,** 196–227.
7. Goss, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46,** 419–428.
8. Goss, A. E. A stimulus-response analysis of the interaction of cue-producing and mediating responses. *Psychol. Rev.*, 1955, **62,** 20–31.
9. Hake, H. W., & Eriksen, C. W. Effect of number of permissible response

categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50,** 161–167.

10. Holton, R. B., & Goss, A. E. The effects of five types of verbalization on acquisition of discriminative motor responses. *J. gen. Psychol.*, 1956, **55,** 117–126.
11. Kurtz, K. H. Discrimination of complex behavior: the relationship of training and test stimuli in transfer of discrimination. *J. exp. Psychol.*, 1955, **50,** 283–292.
12. Lewis, D. Motor skills learning. In *Symposium on psychology of learning basic to military training problems.* Washington, D. C.: Dept. of Defense Research and Development Board, May 1953.
13. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton-Mifflin, 1953.
14. McAllister, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46,** 329–336.
15. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
16. Robinson, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49,** 112–114.
17. Spiker, C. C. Experiments with children on the hypothesis of acquired distinctiveness and equivalence of cues. *Child Developm.*, 1956, **27,** 253–263.
18. Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1955.

(Received March 28, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

RETENTION IN MOTOR LEARNING AS A FUNCTION OF AMOUNT OF PRACTICE AND REST[1]

JOHN C. JAHNKE[2]

Indiana University, Jeffersonville Center

While at least two studies (**2, 6**) of motor learning present information on the combined influence of amount of practice (degree of learning) and length of interpolated rest on short-term retention, the primary purpose of the present study was to examine the effect of these variables on relatively long-term retention. A secondary purpose was to determine the relationship between the slope of the retention curve and degree of learning for the rotary pursuit task.

METHOD

Subjects.—The *Ss* were 240 sophomore, junior, and senior men students attending the Jeffersonville, and New Albany, Indiana, High Schools. All *Ss* were naive to the rotary pursuit apparatus.

Apparatus.—The apparatus was a 60-rpm pursuit rotor which rotated in a clockwise direction and was equipped with a hinged, brass stylus. The center of the 1-in. diameter brass target, which was set flush with the turntable top, was located 3¼ in. from the center of the 11½-in. diameter turntable. Length of work and rest periods was timed manually with a .1-sec. stopwatch.

Procedure.—The experimental design was a 4 × 3 factorial with four levels of practice (either 1, 2½, 5, or 10 min.) with a 15-sec.-work, 45-sec.-rest cycle, and three lengths of interpolated rest (either 10 min., 1 day, or 1 wk.). The 12 groups of 20 *Ss* each thus formed were matched, *S* for *S*, on the basis of total time on target during the first four trials. Following interpolated rest, all *Ss* were given 3 min. of postrest practice, still with the 15–45 cycle. The instructions given to *Ss* were those of a previous study (**7**) with the addition that *Ss* were asked to be seated and relax during each of the rest periods.

RESULTS AND DISCUSSION

Because all groups were matched on initial ability and all received practice with the same work-rest cycle, a graph of prerest performance is not presented here. However, the plot of mean prerest performance for each of the four degree-of-learning conditions disclosed a negatively accelerated growth function with mean terminal prerest trial proficiencies for the four conditions (in order of increasing degree of learning) of 35.31%, 56.81%, 68.54%, and 74.93% time on target. The presentation of averaged data was justified since related-measures analyses of variance performed separately indicated no significant differences in total prerest performance among the three groups within each of the four conditions (all F values < 1, 2 and 38 df).

The data of major concern in the present study are given in Fig. 1, where complete postrest performance for each of the 12 groups is presented, and in Table 1, where percentage time-on-target scores for each of the 12 groups are presented for the last prerest trial, the first postrest trial, and the twelfth (last) postrest trial. One of the salient features of these data is that they suggest that initial postrest performance is related to amount of practice, as might be expected, but is independent of length of rest, even though slight per-

[1] This study was supported in part by a research grant from the Graduate School of Indiana University. The writer is indebted to C. P. Duncan and C. T. Perin, Jr. for critical readings of the manuscript and to Robert Barksdale, Richard Brown, and Larry Sperzel for assistance with the collection of data.

[2] Now at Miami University.

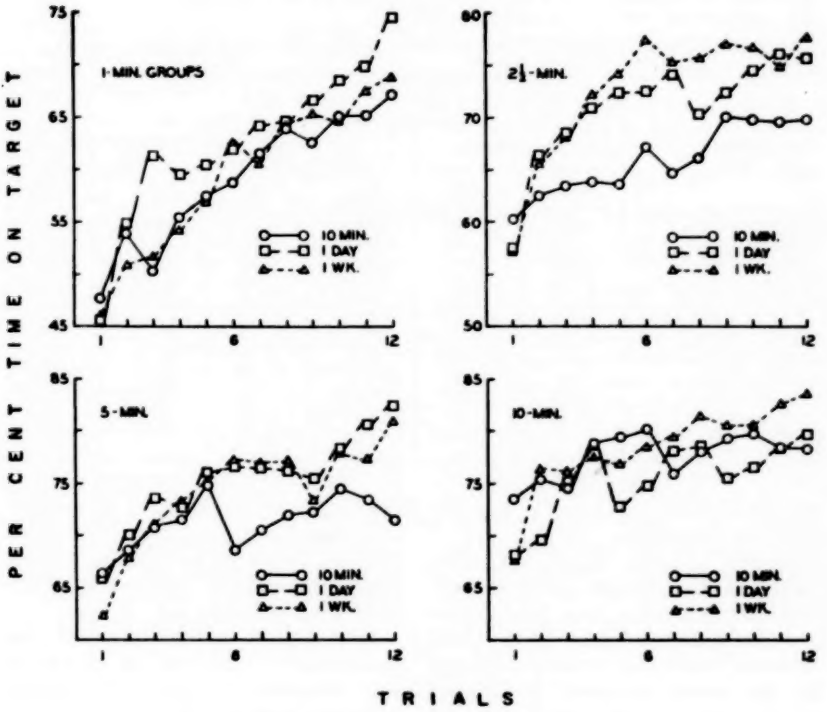

FIG. 1. Postrest performance curves for all groups.

formance losses occurred with increased rest; both suggestions are supported by related-measures analyses of variance. (The F ratios for these variables were 44.47, 3 and 114 df, $P < .01$, and 2.17, 2 and 114 df, $P > .05$, respectively.) The lack of significance of the length of rest variable indicates that the initial trial retention points for each of the four degree-of-learning conditions could be fit by horizontal lines; the slopes of the four retention curves obtained in this manner thus are equivalent and indicate, apparently, no differential rates of forgetting. It should be noted, parenthetically, that the lack of significance of the rest variable stands in opposition to other results (7), and further experimentation seems necessary to reconcile the conflicting findings. Procedural variations perhaps may be held accountable. Another point to be noted is that a comparison of the last prerest trial and first postrest trial scores for each group (see Table 1) discloses that reminiscence gains occurred in the three 1-min. practice groups. Thus, retention curves drawn from the first postrest trial data of the present study seem not to reflect the forgetting process alone; rather, these curves likely represent the confounded effects of forgetting and reminiscence and perhaps are further confounded by an asymptote of performance effect (1, pp. 34–35) in

TABLE 1

PERFORMANCE IN PER CENT TIME ON TARGET FOR ALL GROUPS ON SELECTED TRIALS

Group	Trial					
	Last Prerest		First Postrest		Twelfth Postrest	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
1-Min.—10-Min.	35.15	21.49	47.74	16.57	67.17	13.06
1-Min.—1-Day	35.52	16.82	45.61	18.96	74.64	12.96
1-Min.—1-Wk.	35.16	18.94	45.81	15.05	68.81	16.31
2½-Min.—10-Min.	53.96	22.68	60.24	19.81	69.54	15.19
2½-Min.—1-Day	58.13	18.39	57.70	15.54	75.47	9.23
2½-Min.—1-Wk.	58.36	16.31	57.24	15.58	77.58	10.31
5-Min.—10-Min.	67.24	15.02	66.44	16.45	71.73	11.24
5-Min.—1-Day	66.34	13.20	65.97	12.35	82.54	9.41
5-Min.—1-Wk.	72.02	13.30	62.12	10.63	80.85	10.70
10-Min.—10-Min.	74.78	9.44	73.53	11.50	78.49	12.27
10-Min.—1-Day	74.30	12.79	68.07	14.72	79.80	9.00
10-Min.—1-Wk.	75.70	9.65	67.69	16.38	83.55	8.82

the longer-practice groups and by warm-up decrement.

Several studies (**1, 3, 8**) have suggested that initial postrest performance measures may lead to inferences quite different from those obtained from later postrest practice; for this reason attention is directed to the later postrest trials data of Fig. 1 and Table 1. These data suggest that final postrest performance is positively related to both amount of practice and length of rest; statistical analyses of the twelfth postrest trial data showed, however, that while performance increased reliably with amount of practice ($F = 11.99$, 3 and 114 *df*, $P < .01$), reliable increases with rest occurred only in the 5-min. condition ($F = 6.41$, 2 and 38 *df*, $P < .01$). Increases in later postrest performance with increased rest up to a week and perhaps longer have been obtained previously (**7**), have been taken to indicate the dissipation of an inhibitory process over relatively long periods of time, and have been taken to imply that at least part of the postrest performance decrement previously attributed to a permanent inhibitory process (such as Hull's [**5**] sI_R) may be attributed to undissipated residual temporary inhibition (such as Hull's I_R).

Information on the effect of the independent variables on warm-up was obtained by means of related-measures analyses of variance of the gains in performance from the first to fourth postrest trial shown by each of the 12 groups. (See Fig. 1.) The analyses of these data, which were not corrected for practice effects occurring over the postrest trials (cf. **4,** pp. 359–360), showed that warm-up increased reliably with increased rest in the 2½-min. condition ($F = 4.86$, 2 and 38 *df*, $P < .05$) and that warm-up was not systematically related to degree of learning ($F < 1$, 3 and 114 *df*, $P > .05$). The latter finding opposes other results (**2, 6**), and the reason for this discrepancy is not immediately clear; procedural variations, uncontrolled or unrecognized variables, or unreliability of the phenomenon may all be contributing factors.

Summary

Twelve matched groups of 20 men each were practiced on a pursuit rotor. Each of four sets of three groups received either 1, 2½, 5, or 10 min. of prerest practice under a 15–45-sec. work-rest cycle; one group from each set of three then received either 10 min., 1 day, or 1 wk. of interpolated rest. Three minutes of postrest practice, still with the 15–45-sec. cycle, followed rest for all groups. The essential results were: (*a*) Degree of learning was found to be associated with increases in performance at both initial and final stages of postrest practice. (*b*) Length of rest was not systematically related to initial postrest performance, but during final postrest performance, gains occurred with increased rest up to 1 wk. in length. This result was taken to indicate the rather long-term dissipation of some inhibitory process. (*c*) Warm-up decrement was found to increase with rests greater than 10 min. in only a single degree of learning condition; warm-up decrement was not systematically related to degree of learning.

References

1. Adams, J. A., & Reynolds, B. Effects of shift in distribution of practice conditions following interpolated rest. *J. exp. Psychol.*, 1954, **47,** 32–36.
2. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37,** 393–411.
3. Ammons, R. B. Performance within short practice periods as affected by warm-up decrement and temporary work decrement. *Yearb. Amer. phil. Soc.*, 1955, 181–185.
4. Barch, A. M. Warm-up in massed and distributed pursuit rotor performance. *J. exp. Psychol.*, 1954, **47,** 357–361.
5. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
6. Irion, A. L. Reminiscence in pursuit rotor learning as a function of rest and of amount of prerest practice. *J. exp. Psychol.*, 1949, **39,** 492–499.
7. Jahnke, J. C., & Duncan, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, **52,** 273–282.
8. Reynolds, B., & Adams, J. A. Effect of distribution and shift in distribution of practice within a single training session. *J. exp. Psychol.*, 1953, **46,** 137–145.

(Received April 1, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

ACQUISITION AND EXTINCTION OF PROBLEM-SOLVING SET [1]

R. ALLEN GARDNER [2] AND WILLARD N. RUNQUIST [3]

Northwestern University

The behavioral changes commonly observed in reasoning or problem solving have much in common with those observed in simple associative learning. No experiment or series of experiments can exclude entirely the possibility that qualitatively different processes, such as insight or mental set, distinguish reasoning or problem solving from other classes of behavior. Nevertheless, it is possible to demonstrate the logical and descriptive advantages of analyzing problem-solving behavior by the paradigms of simple associative learning. The experiment described here represents an attempt at such a demonstration.

The Water-Jar Problem, which has been described in detail elsewhere (**2, 3**), was selected as a convenient means of studying typical human problem-solving behavior. Briefly, it is a simple arithmetic problem involving three jars each with a different maximum volume. The task consists of measuring out a required fourth volume by suitable manipulation of the three known volumes. It is possible to construct a series of different problems all of which can be presented in the same form and can be solved by the same steps. Let these steps be called Solution A. If *S* is required to solve a series of problems, which can be solved only by Solution A, then his performance would be expected to improve with each succeeding problem as in the acquisition of other instrumental responses. Similarly, if acquisition of Solution A is followed by extinction to a given criterion, then resistance to extinction would be expected to vary directly with the number of acquisition problems solved. Such an extinction criterion could be provided by a problem of the same kind which could be solved only by a Solution B which is incompatible with Solution A. Assuming that Solution A must be extinguished to some specific level before Solution B is attempted, the time required by *S* to solve a Solution B problem will depend upon the acquisition level of Solution A. Thus, resistance to extinction of Solution A can be measured by performance on a Solution B problem.

Such an analysis of problem-solving set in terms of competing responses is unequivocal in its predictions and economical in its assumptions. It enables the incorporation of problem-solving behavior into a larger body of information. In addition, its assumptions can be made explicit and subjected to direct test. For example, if it is correct to assume that *S* must extinguish Solution A to a given criterion level before he can solve a Solution B problem, then it follows that groups which were at different pre-extinction Solution A performance levels will be equated by

[1] The authors are indebted to Richard P. Youtz of Barnard College who generously provided them with problem materials and much advice on technical details, and made available his unpublished results which in part suggested this experiment.

[2] Now with the Quartermaster Research and Development Center, Natick, Massachusetts.

[3] Now at the State University of Iowa.

forcing them to solve a single Solution B problem. That is, regardless of the differences in Solution A performance created by solving different numbers of Solution A problems, performance on a postextinction Solution A problem will be equal.

Three consequences of the thesis that Water-Jar Problem behavior conforms to a paradigm of simple associative learning can now be stated: (*a*) During acquisition, the time required to solve a problem by a given solution, A, will vary inversely with the number of problems previously solved by Solution A. (*b*) During extinction, the time required to solve a problem by a given solution, B, incompatible with Solution A, will vary directly with the number of problems previously solved by Solution A. (*c*) After extinction to a criterion of one problem solved by Solution B, the time required to solve a Solution A problem will be independent of the number of Solution A problems solved during acquisition.

Method

Problem materials.—Each Water-Jar problem used in this experiment could be solved by only one of two solutions corresponding to Solution A and Solution B in the above discussion. Solution A required filling the largest jar and emptying once into the second largest jar and twice into the smallest jar. It is referred to elsewhere (**1, 2, 3**) as the *b–a–2c* solution, the Einstellung solution, or the indirect solution. Solution B required filling the second largest jar and emptying it once into the smallest jar. It has been referred to as the *a–c* solution or the direct solution.

Procedure.—Undergraduates enrolled in introductory psychology courses at Northwestern University served as *S*s in this experiment. No *S* served in more than one condition. Each *S* was tested individually seated at a table facing *E*. The problems were presented one at a time on 3 × 5-in. cards. Each card bore three numbers labeled *A*, *B*, and *C* representing the given jar volumes, and a fourth number labeled *X* representing the volume to be obtained. A supply of scrap paper was provided for preliminary calculations and *S* indicated each solution by writing it on a separate page of a special pad. Both the problem card and the final solution were removed by *E* after *S* produced a correct solution. Time to solve was measured from the presentation of the problem card by *E* to the completion of the correct solution on the special pad by *S*. Instructions included four practice problems whose solutions were demonstrated by *E*. Among the practice problems were one example of Solution A and one of Solution B. Instructions also included the information that performance would be timed and urged *S* to work as quickly as possible. If *S* gave evidence of not understanding the instructions within the first 2 min. of work on the first problem, the instructions were repeated and the trial was started again.

Experimental conditions.—There were three experimental groups of 36 *S*s each. They received a training series of 6, 11, or 21 Solution A problems, followed by an extinction test which consisted of one Solution B problem. This was followed by a postextinction test which consisted of one Solution A problem. The last training problem, i.e., Problem 6, 11, or 21, was designated as the pre-extinction test. The same two Solution A problems were used for the pre- and postextinction tests. Half of each group received one of these as a pre-extinction test and the other as a postextinction test while this order was reversed for the other half. There were 22 Solution A problems in all. Two of these served as pre- and postextinction problems in the manner described above. The remaining 20 problems were divided into four blocks of five problems. Each block was used as training for an equal number of 6-problem *S*s. For the 11- and 21-problem groups, the blocks were counterbalanced so that each block appeared equally often as the first or last block of training problems. The same Solution B problem was used as the extinction test for all *S*s.

Controls.—There were two control experiments, a Forgetting Control and a Practice Control. The Forgetting Control consisted of two replications of the experimental conditions with 12 *S*s in a group and one alteration in procedure. In place of the extinction problem, these groups were given cartoon books to read between the pre-extinction test and the postextinction test problems (in this case, the pre- and postforgetting problems). One replication read cartoon books for 5 min. and the other for 10 min. The Practice Control consisted of one replication of the 6- and 11-problem conditions with 36 *S*s in a group, and the following alteration in procedure. These groups were

practiced on several simple addition and subtraction problems prior to the training problems in an attempt to compensate for the extra arithmetic practice of the 21-problem group. The 6-problem group was given 45 of these problems, and the 11-problem group, 30. This equalized the actual number of addition and subtraction operations which each *S* had made at the end of the training series.

Results

The time to solve for each problem was recorded to the nearest .1 sec. In order to normalize the highly skewed distributions, raw scores were converted to logarithms. The mean log time to solve for the pre-extinction, extinction, and postextinction tests is plotted for all three groups in Fig. 1. A separate analysis of variance was performed on the log scores for each of the three test problems. For the pre-extinction test, F was 23.33 and for the extinction test F was 5.52. Both are significant with $P < .01$ for 2 and 105 *df*. For the postextinction test F was 2.51, which is not significant ($P = .10$).

The Forgetting Control experiment yielded a measure of the loss in Solution A performance which might be attributed to the time intervening between the pre-extinction test and the postextinction test at the three acquisition levels. The differences were obtained by subtracting time to solve for the pretest from time to solve for the posttest. The raw difference scores were then converted to logarithms for analysis. The means of these log scores appear in Table 1. None of these mean differences is significantly different from zero. The largest t was 1.39 for the 11-problem, 5-min. interval group. Moreover, analysis of variance yielded F's of less than 1.00 for time interval, number of training problems, and interaction. Thus it would seem unreasonable to attribute any significant amount of the postextinction losses of the experimental groups to the time intervening between the pre-extinction test and the postextinction test.

TABLE 1

Mean Log Loss (Sec.) for Groups Given No Extinction Problem

Number of Problems	Time Interval (min.) 5	10
6	−.203	.387
11	.490	.009
21	.256	.329

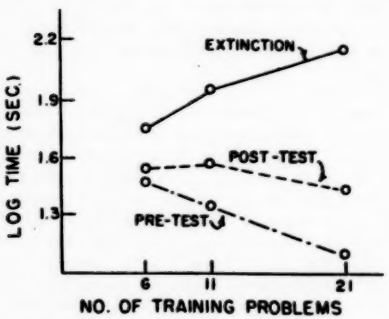

Fig. 1. Mean log time to solve the three test problems as a function of the number of acquisition problems.

The Practice Control also yielded negative results. After equalizing pretraining practice in arithmetic, these groups were indistinguishable from the comparable experimental groups in performance on pre-extinction problems, extinction problems, or postextinction problems.

Discussion

The predictions that the number of training trials would improve training performance and increase resistance to extinction both seem adequately confirmed. While the prediction that performance would be equated by extinction to a criterion was borne out, the confirming data is somewhat less unequivocal. The statistical arguments for

accepting a null hypothesis are seldom as cogent as those for rejecting one, and this prediction must be confirmed by accepting a null hypothesis. Furthermore, the F value did approach significance, and the most deviant point was the superior performance of the 21-problem group which is the one that would be expected to be superior on a common-sense basis in the absence of any specific hypothesis. At the same time, this prediction is more crucial to the theoretical argument in the sense that it is the one least likely to be arrived at by extrapolation of other data or alternative paradigms of problem-solving behavior. Also, it represents an internal check upon the assumption which permits us to use the Solution B problem performance as a measure of extinction to a criterion. The conclusion that extinction to a criterion equated performance is supported by these data more than any other conclusion. The present authors can only suggest further experimentation which provides data for points from 11 problems to 21 problems and beyond in order to determine the stability of the weak trend shown in these data.

There is very little data in the literature of suitable design and technique to offer results comparable to those presented here. The results of an unpublished experiment performed by R. P. Youtz at Barnard College have been made available to the authors by personal communication. As in the present experiment, each problem had only one solution and time to solve was the response measure. Youtz presented 5, 10, 20, and 40 Solution A problems and followed these with 10 Solution B problems. He reports that performance on the first Solution B problem differed significantly, while performance on the second and all subsequent Solution B problems did not. This could indicate that all, or nearly all, of the differential extinction of Solution A takes place during the first Solution B trial which would be in accord with the interpretation presented here.

Kendler, Greenberg, and Richman (1) present an analysis of the Water-Jar Problem behavior and supporting data which suggests that the greater part of problem-solving behavior consists of weakening or extinguishing alternative responses which are incompatible with a required solution. This would seem to agree with the data presented here.

Summary

Three groups ($N = 36$) were trained on 5, 10, or 20 Water-Jar problems which required the same solution. They were then given a pre-extinction test which consisted of one more problem requiring the same solution, an extinction test which consisted of one problem which required a new solution incompatible with the first solution, and a postextinction test which consisted of still another problem which required the first solution.

The results were: (*a*) Performance on the pre-extinction test was a positive function of the number of training problems ($P < .01$). (*b*) Performance on the extinction test was a negative function of the number of training problems ($P < .01$). (*c*) Performance on the post-extinction test was not demonstrably unequal for the three groups ($P = .10$). (*d*) Controls for differential forgetting effects and differential practice effects yielded negative results.

References

1. Kendler, H. H., Greenberg, A., & Richman, H. The influence of massed and distributed practice on the development of mental set. *J. exp. Psychol.*, 1952, **43**, 21–25.
2. Levitt, E. E. The Water-Jar Einstellung Test as a measure of rigidity. *Psychol. Bull.*, 1956, **53**, 347–370.
3. Luchins, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, **54**, No. 6 (Whole No. 248).

(Received April 8, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

CONCEPT FORMATION AS A FUNCTION OF COMPETITION BETWEEN RESPONSE PRODUCED CUES [1]

HOWARD H. KENDLER AND ALAN D. KARASIK

New York University

Human concept formation behavior in a conventional card-sorting test has been conceptualized as involving an implicit mediational mechanism (**5, 6, 7**). The hypothesis was offered that *S* responds to the cards with implicit symbolic responses that serve as cues for his overt card-sorting behavior. Correct card-sorting behavior, according to this formulation, requires *S* to respond to a group of physically different stimuli with a common implicit response.

This mediational analysis was tested by comparing the effectiveness of a reversal shift with a nonreversal shift. It was reasoned that when *S* was making a reversal shift he would, after having learned the first concept, have available the appropriate implicit cue for the second concept. This condition would not exist for a nonreversal *S*. Hence, it was predicted that in a card-sorting test involving two sorting categories a reversal shift would occur more rapidly than a nonreversal shift. This prediction was confirmed in a number of studies (**1, 2, 5, 6**). It is of interest to note that with inarticulate organisms, such as white rats, the opposite relationship prevailed, i.e., a nonreversal shift occurred more rapidly than did a reversal shift (**4**).

The present study seeks to apply this mediational formulation to the problem of verbal concept formation behavior. According to the mediational hypothesis, concept formation occurs when *S* learns to respond to a set of different words with the same implicit response. But this by itself is not sufficient. Not only must *S* respond to the words that belong to a concept (relevant words) with a common implicit response, but *S* must also respond to the words that do not belong (irrelevant words) with different implicit responses. That is, concept formation depends on implicit response differentiation between relevant and irrelevant words.

The implication of this analysis is that concept formation should be influenced by contextual variables. Different sets of irrelevant words should influence the manner in which a set of relevant words are combined into a common concept. Irrelevant words that have strong tendencies to evoke implicit responses that are distinctively different from those required to be given to the relevant words should hasten concept formation. This is the hypothesis that was tested.

METHOD

Subjects.—The *S*s were 72 male and female undergraduates from introductory psychology courses during the 1956 summer session at Washington Square College of Arts and Science, New York University.

Experimental procedure.—The description of the experimental problem is contained in the following experimental instructions read by *E* to the *S*:

"This is a simple experiment that seeks to discover how concepts are formed. You know

[1] This study was supported by the Office of Naval Research. Reproduction of this article in whole or in part is permitted for any purpose of the United States Government.

what a concept is. A concept is a name applied to a group of things which possess one or more common attributes, though they may be quite different in other respects. For example, the nouns *night*, *dungeon*, and *cave* all possess the attributes of being 'dark' and therefore can be considered to belong to a concept that has the common sensory quality of being 'dark.' Your task in this experiment is to form concepts that give a common sensory impression. Another example of a concept with a common sensory impression is one that includes the nouns *silk*, *velvet*, and *skin*. These nouns go together because they all feel 'smooth.'

"In this experiment you will be shown eight cards, each with a different noun typed on it. Four of them will form a concept, in that they have a common sensory property. After looking at the eight cards for 15 sec. you will be given one of the eight cards and asked to select the additional three cards that belong with it—and tell me the common sensory quality on which you have based your selection. Any questions?"

The *S* was given a practice set of eight cards containing the following words: *mansion*, *auditorium*, *city*, *whale*, *pickle*, *pineapple*, *rabbit*, and *rod*. After looking at the cards for 15 sec. *E* handed *S* the card with the word *city* on it, while instructing him to select the three cards from the remaining seven that possessed a common sensory property. The *S* was also instructed to identify the sensory property. If the *S* chose *mansion*, *auditorium*, and *whale*, the *E* said "Right." If *S* did not spontaneously identify the common sensory quality ("big" or its equivalent), he was asked to identify it. If *S* was unable to do so, *E* told him the correct answer. If *S* selected the wrong three cards, *E* said, "Wrong—try again." After two additional errors *E* showed *S* the correct response and explained it to *S*.

The experiment proper was then begun, with the instructions similar to those presented earlier with the practice set. The *S* was given a total of three problems; for each problem *S* was allowed a total of five attempts to select the appropriate set of cards. When a problem was solved, or after five unsuccessful attempts, *E* proceeded to the next problem.

The verbal material used consisted of concrete nouns selected from a list prepared by Underwood and Richardson (**9**). Their list contained 213 concrete nouns that were divided into 40 sense-impression categories. Underwood and Richardson had discovered that each noun had a tendency to evoke one or more sense-impression responses (e.g., round, small). For example, 94% of Underwood and Richardson's *Ss* (introductory psychology students at Northwestern University) responded to the concrete noun stimulus *wheel* with the sense-impression response "round" or its equivalent. The same response was given to the stimulus word *moon* by 30% of their *Ss*, while only 19% of their *Ss* responded in this manner to the stimulus *apple*. Most of the nouns prompted more than one sense-impression response. For example, 67% of the *Ss* responded to the noun *apple* with the sense impression of "red," 19% with the sense impression of "round" and 5% with the sense impression of "sweet."

The variation in the percentages of sense-impression responses to different concrete nouns made it possible to quantify the degree to which each word belonged in a particular category. Thus, it was possible to form groups of words each possessing approximately an equal tendency to prompt a given sense impression. In this experiment high-dominant words were those that evoked a common sense impression in more than 65% of the *Ss*. Medium-dominant words ranged from 25% to 65%, while the range of low-dominant words was between 15% and 28%.

In the present experiment the concepts that were to be formed consisted of four middle-dominant words from a common category. Three such categories were used: "round," "soft," and "small." The words included in each category and their percentages were as follows: (*a*) "round": *platter* (38%), *derby* (33%), *moon* (30%), and *bracelet* (25%); (*b*) "soft": *flannel* (54%), *kitten* (41%), *moccasin* (53%), and *jellyfish* (31%); (*c*) "small": *pill* (46%), *snail* (42%), *minnow* (62%), and *thimble* (37%).

The four words comprising a to-be-formed concept were always presented with four irrelevant words. The set of irrelevant words was presented under three different conditions. In Cond. 1 the set of irrelevant words consisted of four high-dominant words from a common category. The irrelevant words in Cond. 2 were low-dominant words from a common category, while those in Cond. 3 were four high-dominant words selected from four unrelated sense-impression categories.

In order to minimize the possibility that the results would be due to the specific set of irrelevant words used, it was decided to use two different sets of irrelevant words for each of the three conditions. The six sets of irrelevant words are reported in Table 1.

Condition 3 was added to the experimental procedure to isolate the *empirical* variable responsible for any advantage that Cond. 1 might demonstrate over Cond. 2. The set of irrelevant words in Cond. 1 differed from those in Cond. 2 in *two* respects. The irrelevant words

TABLE 1

THE TWO LISTS OF IRRELEVANT WORDS FOR EACH OF THE THREE EXPERIMENTAL CONDITIONS

Experimental Condition								
1 High-Dominant Related			2 Low-Dominant Related			3 High-Dominant Unrelated		
Word	Sensory Impression	%	Word	Sensory Impression	%	Word	Sensory Impression	%
chalk	white	80	enamel	white	28	blood	red	91
milk	white	83	fang	white	10	night	dark	90
snow	white	71	paste	white	16	rhinestone	shiny	67
teeth	white	72	skull	white	22	spinach	green	96
ether	smelly	70	cheese	smelly	25	blush	red	96
garbage	smelly	80	daffodil	smelly	12	dagger	sharp	70
gardenia	smelly	65	goat	smelly	20	eel	slimy	65
manure	smelly	83	sauerkraut	smelly	24	sugar	sweet	82

in Cond. 1 had a higher dominance level to prompt a *common* sense-impression response. But they also had a higher dominance level to evoke *any* sense-impression response. Consider, for example, the low-dominant irrelevant words in Cond. 2 that evoked a "white" sense-impression response. One such word was *skull*. The strongest sense-impression response prompted by the word *skull* was "hard." However, only 36% of the population responded with that sense impression. It will be recalled that all the irrelevant words in Cond. 1 prompted a sense-impression response above the 65% level. Three of the four words in each of the two sets of irrelevant words in Cond. 2 failed to prompt any sense impression above the 65% level. Consequently, any superiority demonstrated by Cond. 1 over Cond. 2 could be due to the words in the former condition possessing a stronger tendency to evoke a *common* sense impression or *any* sense impression. If, however, the performance during Cond. 1 exceeded the performance demonstrated during both Cond. 2 and 3, then it would have to be concluded that the superiority of Cond. 1 over Cond. 2 must be attributed to the greater tendency the former has in prompting a common sense-impression response.

Each *S* was presented with three successive eight-word groups. In each presentation the group consisted of one four-word set of relevant words and one four-word set of irrelevant words from each of the three conditions. Since there was a choice of two different word groups for each irrelevant condition a random selection was made for each *S* with the restriction that the same sensory category would not be used in the high-dominant and low-dominant conditions (i.e., Cond. 1 and 2).

In summary, then, *Ss* were required to select four words from a group of eight words to form a concept. The four relevant words were always of middle-dominance value. The dominance value of the irrelevant words varied from high-dominant related words (Cond. 1) to low-dominant related words (Cond. 2) and finally to high-dominant unrelated words (Cond. 3).

There were six possible orders of presentation of the three groups of relevant words and six possible orders of presentation of the three irrelevant conditions. Consequently there were 36 possible orders of presentation of the combination of the two to form the eight-word group. Each order of presentation was assigned to two *Ss*, thus making the total of 72 *Ss*.

RESULTS

Two response measures were used to evaluate the effects of the three experimental conditions. One measure was the number and percentage of *Ss* making correct sorting responses on the first test trial. The other response measure was the average number of incorrect sortings made throughout the test series. Table 2 reports the results of both response measures.

TABLE 2

CORRECT RESPONSES ON FIRST TEST TRIAL AND INCORRECT RESPONSES DURING TEST SERIES FOR EACH EXPERIMENTAL CONDITION

Experimental Condition	Correct Responses on First Trial		Incorrect Responses During Test Series		
	N	%	Mean	Median	σ
1. High-dominant related words	42	58.3	1.32	0.37	1.93
2. Low-dominant related words	21	29.2	2.53	2.14	2.17
3. High-dominant unrelated words	28	38.9	2.28	1.81	2.21

A χ^2 test of independence indicates that the distribution of right and wrong responses among the three experimental conditions on the first test trial cannot be attributed to chance ($P < .01$). Individual comparisons indicate that the percentage of correct responses in Cond. 1 is significantly different from Cond. 2 ($P < .001$) and Cond. 3 ($.002 > P > .001$). The difference between Cond. 2 and 3 was not significant ($P = .26$).

There were no significant differences among the three conditions in terms of the time required by *S*s on the first trial to make a correct or incorrect response. In each condition, however, correct responses were made significantly more rapidly than incorrect responses.

It will be recalled that each *S* in each of the three conditions was required to make at least five attempts at combining successfully the four relevant words. In eight cases (representing a total of 7 *S*s) the *S* quit in despair before making the five attempts. In evaluating the results for the entire test series, an *S* who failed to make the correct sorting response in five attempts or quit in despair before making five attempts was assigned a score of five errors.

The results for the entire test series are similar to those obtained on the first test trial. An analysis of the differences among conditions using a sign test (**8**) revealed significant differences between Cond. 1 and 2 ($P < .001$) and Cond. 1 and 3 ($P = .05$). The difference between Cond. 2 and 3 was insignificant ($P = .48$).

In all cases except one the *S*, after making a correct sorting response, was able to supply an appropriate concept name to the relevant words. The one exception occurred in Cond. 1. The *S* instead of applying the word "soft" or its equivalent to the relevant words referred to them as "not smelly." In this case the irrelevant words had a high-dominant tendency to evoke the sense-impression response of "smelly."

DISCUSSION

The general idea that instigated the present study was that concept formation behavior involves implicit response differentiation. Successful concept formation does not depend only on *S* making a common implicit response to each of the relevant words. If it did, concept formation would have proceeded at the same rate under all three experimental conditions. The fact

that it did not suggests that the implicit responses to the irrelevant words also influence concept formation behavior.

The superiority of Cond. 1 over Cond. 2 in inducing correct concept formation behavior can be attributed to the irrelevant words in the former condition possessing a stronger tendency to evoke a common implicit response different from that required for the correct grouping of the relevant words. This conclusion is justified considering that Cond. 3, which possessed high-dominant unrelated irrelevant words, was also found to be inferior to Cond. 1.

Although the results obtained under Cond. 3 help to isolate the empirical variable responsible for the superiority of Cond. 1 over Cond. 2, they raise a new explanatory problem. If each of the irrelevant words in Cond. 3 had a high-dominant level to evoke a sense impression different from that required for the relevant words, why should not concept formation proceed as rapidly in Cond. 3 as in Cond. 1? Or to phrase this question differently, why does concept formation occur more readily if the irrelevant words prompt a common sense impression than if they evoke individual sense impressions? According to our analysis, it is not the common sense-impression response that is important but instead the ability of the irrelevant words to prompt distinctively different sense impressions from those evoked by the relevant words. But these two factors in our experimental situation need not be unrelated. In Cond. 1 each of the four irrelevant words has a strong tendency to evoke a common sense impression. It is likely that the probability of each of these irrelevant words *actually* evoking this common sense impression increases by being accompanied by the other three irrelevant words. That is, by a generalization mechanism, the irrelevant words in Cond. 1, as compared with the other two conditions, are more likely to evoke implicit responses distinctively different from the one required for the relevant words.

Why, then, did not the response tendencies of the irrelevant words in Cond. 2 to evoke a common sense impression summate and thereby produce superior performance in Cond. 2 as contrasted with Cond. 3? The present results fail to offer a definitive answer. It may be that the advantage the irrelevant words in Cond. 2 had over Cond. 3 in evoking a common sense impression was offset by the higher dominance values the irrelevant words in Cond. 3 had to prompt any sense-impression response. The important mechanism, according to our analysis, is the ability of the irrelevant words to prompt distinctively different sense impressions. This can be enhanced either by the irrelevant words prompting a common sense impression or by prompting high-dominant sense impressions. Condition 1 was aided by both factors. But Cond. 2 and 3 each possessed only one of these two factors and hence the performance of the *S*s during these two conditions did not differ significantly.

The major implications of this experiment are twofold. Firstly, the results suggest that response differentiation is basic to concept formation. Words are combined into a common concept when they evoke responses distinctively different from responses to words that are irrelevant to that concept. The discrimination behavior of inarticulate organisms is primarily a function of the physical difference between the stimuli to be discriminated, whereas in the present experiment the conceptual discrimination between the relevant and irrelevant words is primarily a function of the differences between the responses each set of words produces.

Secondly, our results contribute to the validity of Underwood and Richardson's scaling technique (**9**). By discovering that the scale values of their verbal material do have an effect on concept formation behavior, one is forced to conclude that their scale has some validity in the sense of Cronbach and Meehl's (**3**) treatment of *construct validity*. This merely means that the valid-

ity of the experimental material scaled by Underwood and Richardson inheres in all of the empirical relationships in which the scale is involved.

Summary

A mediational S–R formulation was applied to verbal concept formation behavior. It was hypothesized that concept formation depended not only on *S* responding to words that belong to a concept (relevant words) with a common implicit response, but also *S* responding to the words that do not belong (irrelevant words) with different implicit responses. It was predicted, therefore, that different sets of irrelevant words should influence the manner in which a set of relevant words are combined into a common concept. This prediction was confirmed. The results suggested that the speed of concept formation varied directly with the ability of the irrelevant words to evoke implicit responses that are distinctively different from those required for the relevant words.

References

1. Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.
2. Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.
3. Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.
4. Kelleher, R. T. Discrimination learning as a function of reversal and nonreversal shifts. *J. exp. Psychol.*, 1956, **51**, 379–384.
5. Kendler, H. H., & D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept-formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.
6. Kendler, H. H., & Mayzner, M. S., Jr. Reversal and nonreversal shifts in card-sorting tests with two or four sorting categories. *J. exp. Psychol.*, 1956, **51**, 244–248.
7. Kendler, H. H., & Vineberg, R. The acquisition of compound concepts as a function of previous training. *J. exp. Psychol.*, 1954, **48**, 252–258.
8. Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.
9. Underwood, B. J., & Richardson, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–96.

(Received April 8, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

TIME-ORDER ERRORS IN COMPARATIVE JUDGMENTS OF HURTFULNESS[1]

ROBERT H. GEERTSMA[2]

University of Chicago

This study served two aims: (*a*) to demonstrate a time-order error (TOE) in comparative judgments of hurtfulness, and (*b*) to investigate the possibility that these judgments may be influenced by anxiety or stress. Since Fechner (**2**) first called attention to such constant errors for lifted weights they have been discovered in many types of judgments, as summarized by Guilford (**3**), and Needham (**12**). However, this effect has not been systematically studied in judgments of hurtfulness. Lacey, Lacey, and Dallenbach (**8**) found a tendency for an overestimation of the second of paired stimuli produced by a needle algometer, but as the length of the intrapair interval was not controlled, and both positive and negative effects were intermixed, their results are not definitive.

The possible relevance of the anxiety-stress dimension is suggested both by clinical observation and experimental work such as that of Moldawsky and Moldawsky (**10**) showing the effects of anxiety on judgmental, attentional, and thinking processes. Anxiety variables were introduced in the present study by means of the Taylor (**14**) Manifest Anxiety Scale (MAS), and by the presentation of stress conditions. This dual procedure was designed to provide for separate and/or interacting effects. The stress conditions closely follow the procedures used by Korchin, Singer, and Ballard (**7**) in their study of the effects of frustration on the reproduction of forms.

METHOD

Apparatus

The method of constant stimulus differences was employed to obtain judgments of the relative hurtfulness of paired electrical stimuli. The *S*s reported whether the comparison stimulus hurt more or less than the standard. A constant current electronic stimulator manufactured by the C. J. Applegate Co. (Model 228) was used as a shocker. Two chromium-plated cuff links were used as electrodes, their square facings serving as contact surfaces. These were amply coated with electrode paste and initially attached with an elastic band to the dorsal and ventral surfaces of *S*'s right arm about 3 in. above the wrist. Midway through the testing the electrodes were switched to the left arm to avoid harmful effects. A Hunter Decade Interval Timer was used to keep the stimulus duration at .5 sec. throughout the experiment. A Cramer three-cam timer controlled the two intrapair intervals, which were 2.5 and 14.5 sec. between the termination of the standard and the onset of the comparison stimulus.

Procedure

Without knowledge of the experiment, 28 night school college students volunteered to serve as *S*s. They were promised and subsequently paid $1.00 per hour for their participation. Previously they had all taken the group form of the Minnesota Multiphasic Personality Inventory. These *S*s represented approximately 75% of the entire female population of two undergraduate psychology classes. Because

[1] This study is part of a thesis submitted to the Department of Psychology of the University of Chicago in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to thank Sheldon J. Korchin, who acted as sponsor, and Donald W. Fiske and Austin H. Riesen for their help and encouragement through all phases of this investigation.

[2] Now at UCLA Medical Center, Department of Psychiatry, Division of Medical Psychology.

Hall and Stride (4) found sex differences for absolute thresholds of reported hurtfulness, only women, who were more numerous in these classes, were asked to participate. On the basis of their MAS scores they were arranged into two groups consisting of the 14 *Ss* scoring below and the 14 scoring above the total group median of 13.5. The mean MAS score for the high scorers was 21.43, and that of the low scorers was 7.14. By random selection half of these high and low scorers were then assigned to each of the stress and neutral conditions, yielding four subgroups, viz. high MAS-neutral, low MAS-neutral, high MAS-stress, and low MAS-stress.

The experimental procedure consisted of three parts administered individually to all *Ss*. Just prior to the first part and between the first and second parts 10 min. of testing for absolute judgments of hurtfulness was done, but those data proved to be unrelated to the present study and will not be reported here. The three parts of the present study were conducted as follows.

Stress and neutral conditions.—The *Ss* in the stress conditions were required to work for 15 min. on a Vigotsky test which had been made unsolvable. Each *S* was told that this was a test of conceptual intelligence, but an easy one solved by almost all high school seniors from a large sample. The *E* then asserted in a confidential manner his belief that *S* would do well, revealing that he had chosen her in order to obtain a sample of achievers. All solutions offered by these *Ss* were rejected with the admonition to keep working toward the correct solution. Finally, after greeting their failure with serious puzzlement *E* observed that the time schedule permitted no delay at this point so that it would be necessary to discuss this failure at the conclusion of the experiment.

The *Ss* assigned to the neutral conditions worked for 15 min. constructing words from the letters of a key word. It was emphasized to them that this activity served the sole purpose of occupying their attention, and the number of words they derived was inconsequential. Through the remainder of the experiment both stress and neutral *Ss* were treated in a business-like though not unfriendly manner.

Testing.—Next, all *Ss* were told that they would be given a number of stimulus pairs, and were asked to judge whether the second stimulus hurt more or less than the first. The electrodes were attached, and six practice trials were given before the experimental series was begun. The intensity of the standard stimulus was an unvarying 3.25 ma. This was paired with each of six comparison stimuli (with intensities of 2.50, 2.80, 3.05, 3.45, 3.70, and 4.00 ma.) 12 times with an intrapair interval of 2.5 sec. and 12 times with an interval of 14.5 sec. The presentations were balanced so that the two intrapair intervals followed one another an unpatterned but equal number of times, and each intensity of the comparison stimulus of the pairs followed each of the six intensities an equal number of times.

Interview.—After testing, the electrodes were removed and a brief interview was conducted to assess *S*'s reactions to the experiment, and to inform the stressed *Ss* of the actual conditions of their failure in a reassuring, anxiety-reducing context. Most *Ss* felt that their sensitivity to pain or ability to differentiate pain was being assessed. Though descriptions of the stimuli varied considerably, all *Ss* agreed that the shocks were unpleasant to some degree. All stressed *Ss* had revealed some discomfort during the stress conditions, and these reactions were continued in the interview. Some of the characteristic behavior emphasized bitter self-depreciation, anger expressed directly toward *E*, and the insistence that a rejected solution was correct. Those *Ss* who in retrospect did not recall feeling badly about their failure were the *Ss* who primarily remembered feeling angry at *E* or the procedure. Two *Ss* who guessed that the purpose of the stress conditions was to arouse their feelings gave testing results in no discernible way dissimilar from those of the rest. It is likely, then, judging from the content and intensity of the stressed *Ss*' responses, that the stress conditions provided a generally effective anxiety-provoking experience.

Analysis

The point of subjective equality (PSE) was determined for each *S*'s distribution of responses for each intrapair interval. Because of variation in the range of these distributions a single method for determining the PSE could not be uniformly applied. The normal interpolation process and the arithmetic mean were used most often. As responses given by the same *Ss* for the two intrapair intervals were to be compared, the analysis of variance used to determine significance levels employed an error term based on *Ss*' pooled interactions.

Results

The response distributions for all *Ss* taken together are plotted in Fig. 1. For the 2.5-sec. intrapair interval the mean PSE is 3.28 ma.,

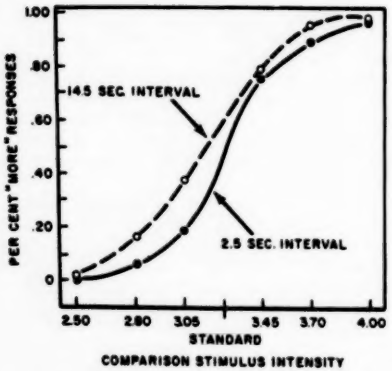

FIG. 1. Distribution of comparative judgments of hurtfulness for the 2.5 and 14.5-sec. intrapair intervals. Each point is based on 72 judgments by each of 28 *Ss*.

whereas for the 14.5-sec. interval it is 3.15 ma. The analysis of variance presented in Table 1 indicates that the difference between these mean PSE values is highly significant.

The interaction between intervals and conditions in Table 1 suggests some influence of the stress conditions on comparative judgments. This effect is made clear in Table 2, which shows a difference between the two conditions only for the 2.5-sec. intrapair interval. This difference between the mean PSE's given by the stress and neutral groups for the shorter interval yields a t of 2.33 ($df = 26$), which is significant beyond the .05 level. For the 2.5-sec. interval, then, the stress group gave significantly *more* judgments than the neutral group of the second stimulus hurting more than the standard.

TABLE 1

ANALYSIS OF VARIANCE OF PSE SCORES

Source	*df*	*MS*	*F*
Conditions (C)	1	.0335	—
MAS levels (MAS)	1	.0002	—
C × MAS	1	.0005	—
Ss in same group	24	.0352	—
Total between *Ss*	27		
Intervals (I)	1	.2327	30.62**
I × C	1	.0297	3.91*
I × MAS	1	.0000	—
I × C × MAS	1	.0086	1.13
Pooled *Ss* × I	24	.0076	—
Total within cells	28		
Total	55		

** $P < .001$.
* $P < .01$.

TABLE 2

MEAN PSE VALUES FOR STRESS AND NEUTRAL GROUPS FOR TWO INTRAPAIR INTERVALS

Condition	*N*	Interval			
		2.5 Sec.		14.5 Sec.	
		Mean	*SD*	Mean	*SD*
Stress	14	3.23	.11	3.15	.12
Neutral	14	3.33	.12	3.15	.18

DISCUSSION

The finding of a positive TOE for the 2.5-sec. intrapair interval and a negative TOE for the 14.5-sec. interval is consistent with almost all comparable work done on other types of comparative judgments. Typically, intrapair intervals less than 3 sec. yield a positive TOE and longer intervals a negative one. Such results have been related to a "fading trace" theory by Kohler (**6**), an "absolute-scale" theory by Muller and Martin (**11**), and Woodrow (**15**), and to a "general adaptation-level" theory by Helson (**5**). In part because of the present finding relating stress to the TOE, though primarily because of its possible heuristic value, a different, admittedly speculative explanation of such effects is offered here.

Comparison of the two stimuli is made difficult by their separation in time. Two events compete for *S*'s attention: (*a*) retention of the impression of the standard, and (*b*) anticipation

of the advent of the comparison stimulus. The resolution of these competing directions of attention may vary with time during the intrapair interval so that for a short interval attention is chiefly directed toward the impression of the standard, whereas for the longer interval attention is increasingly directed toward the comparison stimulus. We assume that attention to the impression of the standard causes an unwitting amplification of that impression, and hence a positive TOE. In support of this assumption the work of Newhall (**13**) and Lalande (**9**) suggests that attention to a tactual image lends it greater emphasis. If the intrapair interval is lengthened, increasing expectancy directed toward the impending comparison stimulus occupies *S*'s attention and causes the comparison stimulus to be overvalued, thus resulting in a negative TOE. The common complaint from *S*s that the long interval causes them to forget the standard could indicate that their expectancy of the comparison stimulus interferes with attention to the impression of the standard

Under the effects of the stress conditions we assume that *S* is less able to attend to the impression of the standard. Evidence from Diethelm and Jones (**1**) suggests that heightened anxiety interferes with attention and concentration. These *S*s then should be primarily involved with the expectancy of the comparison stimulus even for the shorter intrapair interval, resulting in a more negative TOE. For the longer interval it is simplest to suppose that aside from their differential ability to attend to an impression of the standard there are no other relevent effects of the stress.

The above explanation, though speculative, serves to integrate the findings of the present study as well as admitting of some general applicability to TOE effects in other types of judgments. Hopefully its value may lie in presenting the outline for a phenomenologically oriented explanation for the TOE, and an explanation not simply relevant to this particular judgmental phenomenon.

Summary

The method of constant stimulus differences was used to elicit judgments of hurtfulness from 28 female *S*s in response to electrical stimuli. The dimension of anxiety was introduced by means of the Taylor MAS and imposed stress conditions. The stressed group was subjected to failure on what was presented to them as an easy test of conceptual intelligence, and a neutral group occupied themselves for an equal period of time with a word game. All *S*s were then presented with stimulus pairs separated by intervals of 2.5 or 14.5 sec., and required to judge relative hurtfulness.

The results indicate that: (*a*) There is a positive TOE for judgments of hurtfulness with a short intrapair interval, and a negative TOE for a long interval; (*b*) stress conditions systematically influence comparative judgments for the short intrapair interval, resulting in a more negative TOE.

These findings were discussed in terms of an attentional theory of the judgmental process.

References

1. Diethelm, O., & Jones, M. R. Influence of anxiety on attention, learning, retention and thinking. *Arch. Neurol. Psychiat.*, 1947, **58,** 325–336.
2. Fechner, G. T. *Elemente der Psychophysik.* Leipzig, 1869.
3. Guilford, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.
4. Hall, K. R. L., & Stride, E. The varying response to pain in psychiatric disorders: a study in abnormal psychology. *Brit. J. med. Psychol.*, 1954, **27,** 48–60.
5. Helson, H. Adaptation-level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60,** 1-29.
6. Kohler, W. Zur Theorie des Sukessivergleichs und der Zeitfehler. *Psychol. Forsch.*, 1923, **4,** 115–175.
7. Korchin, S. J., Singer, J. L., & Ballard, R. G. The influence of frustration on the reproduction of visually-perceived forms. *J. Pers.*, 1951, **1,** 54–66.
8. Lacey, J., Lacey, B., & Dallenbach, K. Areal and temporal variations in pain sensitivity. *Amer. J. Psychol.*, 1941, **54,** 413–417.
9. Lalande, A. Sur un effet particulier de l'attention appliquee aux images. *Rev. Phil.*, 1893, **35,** 285–287.

10. Moldawsky, S., & Moldawsky, P. C. Digit span as an anxiety indicator. *J. consult. Psychol.*, 1952, **16,** 115–118.
11. Muller, G. E., & Martin, L. J. *Zur Analyse der Unterschiedsemfindlichkeit.* Leipzig, 1899.
12. Needham, J. G. The time-error in comparison judgments. *Psychol. Bull.*, 1934, **31,** 229–243.
13. Newhall, S. M. Effects of attention on the intensity of cutaneous pressure and on visual brightness. *Arch. Psychol., N.Y.*, 1923, No. 61.
14. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48,** 285–290.
15. Woodrow, H. Weight discrimination with a varying standard. *Amer. J. Psychol.*, 1933, **45,** 391–416.

(Received April 12, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

THE INFLUENCE OF COMPLEXITY AND NOVELTY IN VISUAL FIGURES ON ORIENTING RESPONSES[1]

D. E. BERLYNE

Center for Advanced Study in the Behavioral Sciences[2]

The present experiments were meant to continue a line of investigation initiated in a previous series (**10**). In the earlier experiments, undergraduate *Ss* performed key-pressing responses which produced tachistoscopic exposures of visual figures. They were free to give themselves as many exposures of each figure as they wished before the next figure was made available, and the number of responses was regarded as a measure of the "perceptual curiosity" aroused by each figure. It was found, in accordance with predictions from a theory of human curiosity (**7, 8**), that incongruous and surprising figures, which, it was postulated, would give rise to conflict between implicit responses, were more curiosity-arousing than others. A positive relation also emerged between perceptual curiosity and certain other attributes of figures, which can jointly be described by some such term as *complexity*. Some of these attributes, equivalent to the amount of variety in a figure, were ones that Attneave (**2**) has identified with the information-theory measures, redundancy and relative uncertainty (=1−redundancy). Another was *amount of material*, which can be associated with (absolute) uncertainty or with measures of structural information (**15**).

The aim of the present experiments was to supplement these findings by determining whether variables that affect perceptual curiosity also affect attention. The word "attention" has had a variety of meanings in psychology (**4**). In a wide sense, it refers to the problem of selective response in general: an organism is usually surrounded by numerous stimulus objects with the power to evoke responses, and the factors favoring response to one stimulus rather than another need to be identified. In a narrower sense, attention refers to the selective evocation of observing responses (**21**) in particular, i.e., responses that expose the organism to stimuli or aspects of stimuli that were not accessible before. A still narrower sense, which is pertinent to the present inquiry, refers to the direction of observing responses of a special sort, namely, receptor-orienting movements. In the previous experiments, pressing the key was an observing response, but only one figure was available at once, so that there was no competition between stimuli for attention. In the present experiments, two figures were presented side by side on a screen for 10 sec., and eye movements were recorded. Which figure *S* looked at first and how much time he spent looking at each were the data of interest.

The two experiments tested the influence on attention of two kinds of factors that have been shown to

[1] The author is greatly indebted to Leonard Hommel, who acted as assistant in preparing for the experiments and as second *E* during the collection of the data, as well as to Olga McNemar and D. H. Lawrence, who kindly arranged for *Ss* and facilities at the Psychology Department of Stanford University.

[2] On leave of absence from the University of Aberdeen, Scotland. Now at the University of California, Berkeley.

influence perceptual curiosity. Experiment I was concerned with the complexity or information-theory factors that were studied in the previous project. These factors were to be examined more closely by isolating a number of more precisely definable variables and investigating their effects separately.

Experiment II was devoted to a second factor, which has often been found conducive to perceptual curiosity in lower animals, namely, stimulus change or relative novelty. Rats, for example, will spend more time sniffing at an object that was not present a short time ago than at one that was (**3**). In human beings, change seems to draw attention in the widest sense: when given a free choice between pressing a key corresponding to a recently changed stimulus and pressing another key corresponding to a stimulus that has remained unchanged, *S*s show a preference for the former response (**5**). Experiment II investigated further this apparent link between perceptual curiosity and attention by testing the hypothesis that changing features of the environment will come to attract more visual fixation than recurrent features. For a number of training trials, the same figure kept on reappearing on one side of the screen, while the other side bore a different figure every time. Three test trials, in each of which new figures appeared on both sides, were then given to ascertain whether any tendency to concentrate attention away from the side of the constant stimulus persisted after the constant stimulus was superseded.

Method

Procedure

The *S* sat facing a cardboard screen at a distance of about 4 ft. He was told: "This is an experiment on number estimation. We want to find out what factors determine how well people can estimate number. Some pictures will be shown on the screen, two at a time. When you hear the word 'Ready!' you are asked to fixate the cross in the middle of the screen. When the pictures appear, you can look at them if you like, but we should like you to be looking at the cross just before they appear. At the end, you will be asked to estimate how many pairs of pictures there were."

The stimulus material was then presented. Two figures at a time were projected on the screen by means of an overhead projector, operated by the first *E*. Each pair of figures was visible for 10 sec.; there was an interval of 20 sec. between successive pairs, and the warning was given 3 sec. before the figures appeared. The slides were projected as 6×4-in. rectangles, 12 in. apart. The fixation cross was midway between them.

The second *E*, sitting behind the screen, observed *S* through a hole situated below the level at which the figures appeared. He recorded the time spent looking at each figure in units of .2 sec. by means of two Morse keys connected with stop clocks. He also noted which figure in each pair was fixated first.

This way of timing fixations was not infallibly accurate, but any errors were prevented from becoming constant errors by the design of the experiment and by the fact that the *E* responsible for the recording was kept ignorant of the experimental conditions allotted to each *S* and thus of the nature of the figure presented on a particular side during a particular trial. The two *E*s alternated between the tasks of observing eye movements and operating the projector. Although the instructions were designed to make any fixation of a figure as spontaneous as possible by disguising the purpose of the experiment, the recording apparatus made an audible click with each make and break, and *S*s invariably realized after a few trials that their eye movements were being registered.

The material for Exp. II was presented immediately after the material for Exp. I without any interruption of the procedure or warning.

Subjects

Ten male and 10 female undergraduates took part in both experiments.

Stimulus Material

Experiment I.—Twenty-two pairs of slides were prepared, falling into six categories. Categories A, B, C, D and E contained four

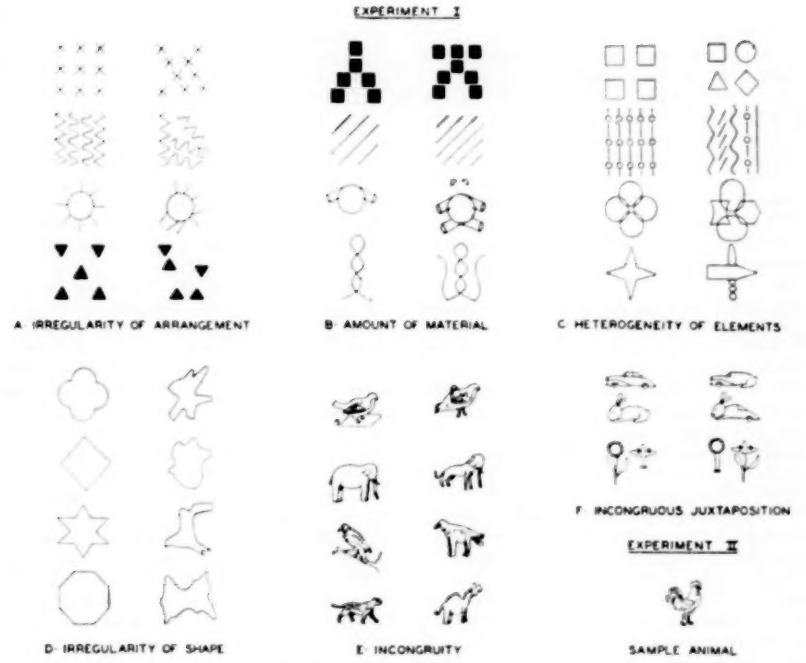

FIG. 1. The stimulus material.

pairs each, and Category F the remaining two pairs (Fig. 1). One member of each pair of slides bore a "less complex" and the other a "more complex" figure, and different forms of complexity were represented by the different categories of material:

A. Irregularity of arrangement.—The two figures in each pair consisted of exactly the same elements, but in the one case they were arranged in a regular geometrical pattern, and in the other they were irregularly scattered. As far as spatial location is concerned, there is more redundancy in the former. A view of part of the figure would enable the probable arrangement of the remaining material to be guessed, given the assumption, based on past experience, that one part of a visual pattern is more likely than not to resemble another.

B. Amount of material.—One figure in each pair consisted of part of the material in the accompanying figure. The latter can be considered to belong to a class of figures with more absolute uncertainty, since more material implies greater possibilities of variation and therefore more information, on the average, to be absorbed before a member of the class can be identified. Alternatively, and perhaps more fruitfully, the concept of structural information content, introduced by MacKay (**15**) after Gabor and supplementing Shannon's theory of selective information (**20**), might be applied. Figures with more elements contain more "distinguishable groups or clusters" or more degrees of freedom and hence more "logons."

C. Heterogeneity of elements.—One figure in each pair consisted of a number of identical elements, whereas the other figure consisted of the same number of different elements in a similar spatial arrangement. Redundancy is an appropriate concept here as in Category A.

D. Irregularity of shape.—Each pair contained one regular geometrical shape, symmetrical along two or more axes, and one irregular nonsymmetrical shape. The variable involved here is one whose relation to redundancy is discussed at length by Attneave (**2**). It may be said, in brief, that regularity of contour and symmetry imply similarity of parts

and predictability of changes in curvature, all of which means high redundancy.

E. Incongruity.—Each pair consisted of a picture of a normal animal and a picture of an incongruous animal, i.e., one with parts appropriate to different species or with three heads. The latter thus possessed attributes which *Ss*' past training would have led them to regard as incompatible and which would consequently lead to conceptual conflict (**7**). The perception of one part could similarly be assumed to conflict with expectations or redintegrative responses aroused by other parts (**6**). The difference between the pictures in each pair could also be expressed in information-theory terms. A picture of a well known animal is highly redundant, as the sight of one characteristic feature makes it easy to reconstruct the rest. One feature of an incongruous animal, on the other hand, gives little hint of what the rest will be like. Each feature has a high information content, since the probability of its appearing together with the other features is low.

F. Incongruous juxtaposition.—This category was added as a control for Category E. If it were found that the incongruous animals attracted more attention than the others, it would still not be clear whether the decisive factor was merely the unusual combination of elements or whether the spatial arrangement also played a part, e.g., whether it was a matter of seeing a bird's head and an elephant's forelegs in unwonted proximity or a matter of seeing a bird's head where an elephant's head would be expected. Each pair belonging to Category F had, therefore, one slide bearing adjacent pictures of two objects not usually seen together and one slide bearing the same material but with the halves of the two objects incongruously juxtaposed.

The 20 pairs of figures in Categories A-E were arranged in a sequence, such that each successive set of five comprised one pair from each category in a randomized order. The two pairs of Category F were then inserted into the sequence in randomly selected positions. For half the pairs in each category, the "more complex" figure was assigned an odd number and the "less complex" an even number; the reverse held for the remaining half.

The 20 *Ss* were divided into five subgroups of 4. One *S* in each subgroup received the material in the original order with odd-numbered figures on the right, one in the original order with odd-numbered figures on the left, one in the original order reversed with odd-numbered figures on the right, and one in the reversed order with odd-numbered figures on the left. Two subgroups contained men only and two contained women only. The fifth subgroup contained two men and two women, the men taking the original-order odd-right and the reversed-order odd-left conditions.

Experiment II.—Seventeen slides bearing pictures of animals were prepared and arranged in a random order. Slides 1–11 were designated as training slides and Slides 12–17 as test slides. Whereas every figure used in Exp. I had a rectangular white background, those used in Exp. II had less extensive white backgrounds leaving only narrow margins around their contours.

Half of the *Ss* received 10 training trials and the remainder 5 training trials. Each of the slides numbered 1–5 served as a recurring stimulus for a different subgroup of 4 *Ss*. It was presented on one side of the screen during every training trial. Other training slides varying from trial to trial, appeared on the other side. In each subgroup, the conditions 10R, 10L, 5R, and 5L were allotted to one *S* each, where 10 and 5 represent the number of training trials and L and R the side on which the recurring stimulus appeared. When there were 10 training trials, the varying stimuli were Slides 1–11 except for the recurring picture. When there were five, the varying stimuli were Slides 7–11. The varying stimuli always appeared in numerical order. Immediately after the training trials, Slides 12–17 were projected, two at a time, for the three test trials. They appeared in the same order and locations for all *Ss*.

Results

First Fixations

Which figure of a pair was fixated first bore no consistent relation to the experimental variables. The *Ss* generally began with the left-hand figure about twice as often as with the right-hand figure, a consequence, presumably, of reading habits. What follows is therefore confined to the other dependent variable, i.e., the amount of time spent looking at each figure in the course of the 10-sec. exposure.

Fixation Times

Experiment I.—The mean fixation time for each of the 44 figures was calculated, as well as the over-all means for "more complex" and "less

TABLE 1

MEAN FIXATION TIMES IN SECONDS: EXP. I

Category	Variable	"Less Complex" Figures	"More Complex" Figures	Wilcoxon's T	P
A	Irregularity of arrangement	3.36	4.56	22	<.01
B	Amount of material	3.40	4.29	28	<.01
C	Heterogeneity of elements	3.16	4.95	7	<.01
D	Irregularity of shape	3.05	4.67	8	<.01
E	Incongruity	3.46	4.97	6	<.01
F	Incongruous juxtaposition	3.46	4.83	14	<.01

complex" figures in each category. The latter are displayed in Table 1. Every single pair showed a greater mean fixation time for the "more complex" figure than for the "less complex" figure. The over-all means for each category were subjected to Wilcoxon's T test (**17**) for paired replicates. The tendency to spend more time on "more complex" figures was signficant beyond the .01 level with every category.

Experiment II.—Figure 2 shows the mean fixation times for varying and recurring figures during the training trials of Exp. II. It will be seen that the 10-trial *S*s spend more and more time on the former and less and less on the latter. The decline in fixation time for the recurring stimuli was subjected to an analysis of variance for trend (**1**). Since the fixation times for recurring and varying stimuli were necessarily interdependent, there was no point in analyzing both. Both the linear trend and the deviation from linearity turned out to be significant, as Table 2 makes clear. The 5-trial group had exactly the same stimuli during their training trials as the 10-trial group had during their last five training trials. Their curves show a similar divergence but, as was to be expected, are not so far apart as the corresponding portions of the curves for the 10-

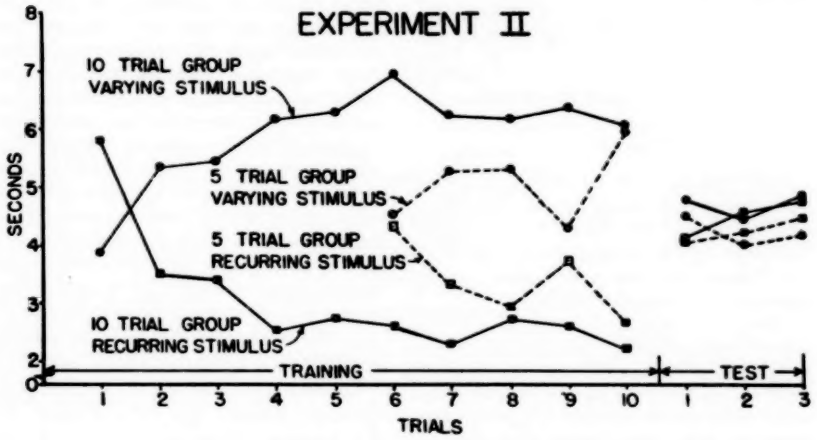

FIG. 2. Fixation times in Exp. II.

TABLE 2

ANALYSIS OF VARIANCE OF FIXATION TIMES FOR RECURRING STIMULUS IN TRAINING TRIALS, 10-TRIAL *Ss*, EXP. II

Source	*df*	*MS*	*F*	*P*	Error Term
A. Trials	9	11.23	6.93	<.001	E
A1. Linear trend	1	56.11	11.74	<.01	E1
A2. Deviation from linearity	8	5.62	4.57	<.001	E2
B. Sexes	1	2.79			D
C. Sexes × Trials	9	0.71			E
D. Individuals within sexes	8	7.65	4.72	<.001	E
E. Individuals × Trials	72	1.62			
E1. Individual linear trends	8	4.78			
E2. Individual deviations from linearity	64	1.23			
Total	99				

trial group. An analysis of variance for trend was carried out on the recurring-stimulus data for the 5-trial group but failed to produce any significant *F* values. Moreover, the difference between the mean fixation times for recurring stimuli on Trials 1–5 for the 5-trial group and Trials 6–10 for the 10-trial group did not turn out to be significant according to a *t* test, although it was in the expected direction, the means being 3.37 and 2.50 respectively.

As for the test trials, the graph shows that *Ss* reverted to fixating both sides about equally. There was no consistent tendency to devote more time to the side formerly occupied by the varying stimuli.

DISCUSSION

The findings can be discussed either in information-theory or in behavior-theory terms, the field of exploratory behavior and observing responses being one where the two approaches can be expected to converge (**11**).

When *S* is faced with a definite discriminative task, such as naming a stimulus or pressing an appropriate key as quickly as possible, it has regularly been observed that reaction time increases with information-theory measures such as the initial uncertainty or the amount of transmitted information (**9**). And when letter sequences are exposed in brief flashes of constant duration, the likelihood of misidentification increases with information content (**16**). These facts have often been taken to mean that the human being is a communication channel with a limited capacity and requires more time to absorb more information. In the light of such an assumption, it may not be surprising that *Ss* spent more time on the inspection of more complex or conflict-arousing, i.e., more uncertainty-laden, stimuli in the preceding series of experiments with the tachistoscope (**10**). What is interesting, however, is that *Ss* of their own accord spent more time looking at the more uncertainty-laden figures, when no specific task requiring identification of the figures was imposed on them and any persistence in absorbing the information completely was spontaneous. The present experiments illustrate quite a different, if related, point, which does not necessarily follow at all from the postulate that the rate of taking in information is limited; they reveal a tendency to fixate a part of the environment that is a relatively rich source of information in preference to one that is a relatively poor one.

To turn from information theory to behavior theory, the data yielded by the present experiments do not by themselves call for any such intervening

variable as "curiosity" or "exploratory drive." Since we have a movement focusing a sense organ on a stimulus object that has already been perceived peripherally, the investigatory-reflex theory of Pavlov (**19**) provides a parsimonious explanation. We should have to conclude that "more complex" stimuli elicit the investigatory reflex more strongly than others, as well as that new stimuli elicit it more than stimuli that have been present for some time, which is a fact frequently noted by Pavlov and his school. In the preceding set of experiments (**10**), on the other hand, the response was an arbitrary one, namely, pressing a key, and the stimulus was not visible at all until the response had been performed. The probability of repeating the response can thus be attributed to an intervening variable, which can be called "perceptual curiosity." The strength of the perceptual curiosity aroused by different figures, judged by the number of times *S*s exposed themselves to them, increased with the same "complexity" variables as were found to influence attention in Exp. I of the present series. Experiment II showed that more attention will be given to stimuli that have newly come on to the scene, just as lower animals have been found repeatedly (e.g., **3**) to devote more exploration to them. By verifying that eye movements are influenced by two sorts of variable that apparently affect perceptual curiosity, we have confirmed that attention in the narrowest sense depends on, among other things, the curiosity-arousing properties of stimuli. This is in line with a theory of attention in the widest sense that was outlined in an earlier article (**4**). A further hypothesis (**7, 8, 11**)—that complex or novel stimuli attract observing responses and other investigatory behavior because incomplete perception of them arouses a drive which continued examination of them reduces—is therefore admissible and deserving of consideration.

Of the "more complex" stimuli in Exp. I, those in Categories E (incongruity) and F (incongruous juxtaposition) can plausibly be held to produce *conceptual conflict.* Since the forms of complexity represented by the other categories have evidently much the same affects on curiosity and attention, since they, like incongruity, admit of an information-theory analysis, and since information-theory measures, as used in psychology, seem to relate to conflict situations, the possibility that all forms of stimulus complexity involve conflict is worth entertaining (**11**). The conflict could be between various incompatible perceptual responses (**4**), e.g., between the classifying responses or "connotative meanings" (**18**) evoked by different parts or aspects of a pattern, or between the perception of one part or aspect and an expectation aroused by another (**6, 14**).

It may be objected that "less complex" figures are likely to resemble familiar stimulus patterns to a greater degree than "more complex" ones, and that this difference in novelty could be responsible for our results. This point is well taken. But the suggestion (**7**) that stimulus complexes evoke conflict to the extent that they are complex or unfamiliar is offered as a possible explanation of the fact, among others, that new stimuli and well known stimuli affect organisms differently.

In conclusion, some limitations of the present study must be indicated. One is that the "more complex" figures were not tremendously complex. Much more complex patterns might conceivably have been shunned rather than preferentially inspected. Aestheticians have often asserted that an intermediate degree of complexity ("unity in diversity") makes for maximum appeal, the exact degree depending on personality traits (**13**). Too little complexity means insipidity and too much means bewilderment.

Shortening or lengthening the time of exposure might have brought about different results.[3] The failure of the

[3] In a more recent experiment with 2-min. exposures, 51% of the time was spent fixating "more complex" figures and 36% fixating "less complex" figures.

independent variables to affect the direction of the initial fixation was probably due to the fact that, with a 10-sec. exposure, *Ss* knew they had ample time to examine both figures, no matter with which they started. A brief flash, such as Brandt (12) used, would have allowed no more than one figure in each pair to be fixated, and differences in power to attract attention or arouse perceptual curiosity might have overcome reading habits.

Summary

Undergraduate *Ss* were presented with a succession of pairs of visual figures, projected on a screen for 10 sec. each. Which figure was fixated first and how long each member of a pair was fixated were the data recorded. In Exp. I, one figure of each pair was "more complex" or, in information-theory language, higher in relative or absolute uncertainty than the other. Six different variables which may be regarded as forms of complexity were investigated separately. In all cases, significantly more time was spent looking at the "more complex" figure. In Exp. II, the same figure recurred on one side of the screen for a number of training trials, while a new figure appeared every time on the other side. The fixation time for the varying stimuli progressively increased at the expense of the fixation time for the recurring stimulus. During subsequent test trials, when new material appeared on both sides, *Ss* reverted to fixating both figures about equally. Which figure was fixated first bore no consistent relation to the independent variables.

The implications of the findings are discussed in terms of psychological information theory and of behavior theory, with special reference to the relation between attention and perceptual curiosity.

References

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43,** 533–557.
2. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61,** 183–193.
3. Berlyne, D. E. Novelty and curiosity as determinants of exploratory behavior. *Brit. J. Psychol.*, 1950, **41,** 68–80.
4. Berlyne, D. E. Attention, perception and behavior theory. *Psychol. Rev.*, 1951, **58,** 137–146.
5. Berlyne, D. E. Attention to change. *Brit. J. Psychol.*, 1951, **42,** 269–278.
6. Berlyne, D. E. Knowledge and stimulus-response psychology. *Psychol. Rev.*, 1954, **61,** 245–254.
7. Berlyne, D. E. A theory of human curiosity. *Brit. J. Psychol.*, 1954. **45** 180–191.
8. Berlyne, D. E. An experimental study of human curiosity. *Brit. J. Psychol.*, 1954, **45,** 256–265.
9. Berlyne, D. E. Conflict and choice time. *Brit. J. Psychol.*, 1957, **48,** 106–118.
10. Berlyne, D. E. Conflict and information-theory variables as determinants of human perceptual curiosity. *J. exp. Psychol.*, 1957, **53,** 399–404.
11. Berlyne, D. E. Uncertainty and conflict: a point of contact between information-theory and behavior-theory concepts. *Psychol. Rev.*, 1957, **64,** 329–339.
12. Brandt, H. F. *The science of seeing.* New York: Philosophical Library, 1944.
13. Eysenck, H. J. The experimental study of the 'good Gestalt'—a new approach. *Psychol. Rev.*, 1942, **49,** 344–364.
14. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
15. MacKay, D. M. Quantal aspects of scientific information. *Phil. Mag.*, 1950, **41,** 289–311.
16. Miller, G. A., Bruner, J. S., & Postman, L. Familiarity of letter sequences and tachistoscopic identification. *J. genet. Psychol.*, 1954, **50,** 129–139.
17. Mosteller F., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954.
18. Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49,** 197–237.
19. Pavlov, I. P. *Conditioned reflexes.* Oxford: Oxford Univer. Press, 1927.
20. Shannon, C. E., & Weaver W. *The mathematical theory of communication.* Urbana: Univer. of Ill. Press, 1949.
21. Wyckoff, L. B. The role of observing responses in discrimination learning. Part I. *Psychol. Rev.*, 1952, **59,** 431–442.

(Received April 12, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

INTERACTION BETWEEN TOTAL STIMULUS INFORMATION AND SPECIFIC STIMULUS INFORMATION IN VISUAL RECOGNITION [1]

J. R. NEWBROUGH

University of Utah

It has been generally accepted that the information measure will yield reliable predictive results when used in psychological experimentation concerned with visual recognition. That is, various authors have found that psychophysical performance is a strictly monotonic function of the measure of uncertainty. To illustrate, Archer (**1**) found that when various patterns were presented on an oscilloscope, *S*s' reaction times in correctly classifying them were a linear function of the total number of alternatives or total stimulus information (I_{ts}). Frick (**3**) showed that recognition time to words presented on a screen increased linearly with the number of options available (I_{ts}). These are two of many experiments reporting similar results which seem to substantiate this already accepted application of the information measure.

The results reported by Hyman (**4**) suggest that this application of the Information Theory model to psychophysical phenomena needs further study. He substantiated the findings of Archer (**1**) and Frick (**3**) by demonstrating that reaction time to differing patterns of lights increased linearly with an increase of the I_{ts}. But, more important, he also noted that it was impossible to predict reaction time to individual stimuli from the specific stimulus information (I_{ss}) values along the function obtained for I_{ts}. The reaction time to the low information stimuli was found to be higher, and to the high information stimuli lower, than would be predicted from the regression line. To explain these unpredicted results he suggested the possibility of an interaction effect among the various stimuli within an experimental condition.

The purpose of this study is to examine the effect of stimulus interaction upon perceptual behavior. There are two ways in which the interaction could affect performance: (*a*) by the interaction of the I_{ss} values among themselves (as found by Hyman), and/or (*b*) by the interaction of the I_{ts} and the I_{ss} values. The I_{ss} interaction effect is to be evaluated by keeping the I_{ts} values constant, while the internal structure (I_{ss} values) varies. If this interaction be present, significant differences in responses should be obtained. The I_{ts}-I_{ss} interaction is to be examined by holding a I_{ss} value constant across several conditions which vary as to I_{ts} values. This interaction should be manifest by significantly different responses to the constant I_{ss} value.

METHOD

Subjects.—The *S*s were 36 men and women students in introductory psychology courses. They performed in three groups of 12 each.

[1] This article is based upon a thesis submitted in partial fulfillment of the requirements for the M.A. degree at University of Utah, 1956. The writer wishes to express grateful acknowledgment to J. L. Morey for his generous assistance and helpful direction of the research, and to P. B. Porter and L. E. Bourne, Jr., for their aid in the preparation of this manuscript.

Apparatus.—The apparatus consisted of a tachistoscope, an opaque projector fitted with a gear-driven shutter and iris diaphragm, which flashed the words on a white screen. The room was darkened to the point where the words were easily visible on the screen, and yet *S*s could see to write on the data sheets. The ambient light on the screen was equated for all three groups.

The projector was equipped with a white light pointer which served as a fixation point and the words appeared centered below it. The words were printed in 18-point lower-case black type on cards of heavy white poster board. These were inserted one at a time into the projector. The letters appeared 1½ to 2 in. high on the screen.

Stimulus material.—The seven words *mean*, *mine*, *more*, *nice*, *rain*, *iron*, and *wave* were selected from the Thorndike-Lorge word counts (6) as being within the first 1000 most frequently appearing words, composed of only four letters which did not project above or below the body of the word, and maximally confused with each other as determined by preliminary experimentation.

Table 1 illustrates the stimulus material as it was fitted into series for presentation. There were five series of 40 randomly ordered presentations each. The number of words in each series was varied from 2 in Series A to 7 in Series D. The probability values appearing in the column to the far right indicate the number of times each word appeared in the series. For example, in Series A, *mean* had a probability value of .25 indicating that it occurred 10 times in the 40 presentations.

It should be noted that Series B, C, and D have essentially the same I_{ts} value (2.00 bits) but that the I_{ss} values of the words within each series are different. Also, the word *mean* had the same I_{ss} value (2.00 bits) in Series A, B, D, and E.

The words for each stimulus series and their respective probabilities of occurrence in percentages were lettered 1½ in. high with white paint on black poster paper. Before each series was presented, the poster displaying its structure was hung to the left of the screen. This was done to convey the I_{ts} and I_{ss} values to *S*s through the probability values.

Scoring.—The dependent variable was a *missed score* obtained by dividing the number of presentations missed (recorded as "not seen") by the total number of presentations. A recognition score (number seen divided by total number of presentations) was found in preliminary experimentation to be inapplicable because of its greater dependence upon the response preferences of *S*s.

TABLE 1

THE FIVE SERIES AND THE VALUES OF I_{ts} AND I_{ss} IN BITS FOR EACH WORD IN THE SERIES

Series	Number of Alternatives	I_{ts}	Words	I_{ss}	Prob. of Occurrence
A	2	.81	*mean**	2.00	.25
			mine	.41	.75
B	4	1.96	*mean*	2.00	.25
			nice	2.32	.20
			more	2.32	.20
			mine	1.51	.35
C	5	2.00	nice	3.00	.125
			more	3.00	.125
			rain	3.00	.125
			mean	3.00	.125
			mine	1.00	.50
D	4	2.00	*mean*	2.00	.25
			nice	2.00	.25
			more	2.00	.25
			mine	2.00	.25
E	7	2.75	*mean*	2.00	.25
			mine	3.00	.125
			more	3.00	.125
			nice	3.00	.125
			rain	3.00	.125
			iron	3.00	.125
			wave	3.00	.125

* *Mean* has the common I_{ss} value (2.00 bits) in Series A, B, D, E.

Experimental procedure.—The *S*s were seated in the three arching rows of chairs 14 to 21 ft. from the screen and given preliminary instructions about what to expect, and how to record what was seen. It was particularly stressed that a word was to be recorded only if *S* was certain of its identity; otherwise a question mark. The poster to the left of the screen would be changed before each series and would tell them what words would be contained therein. The fixation point was explained and a trial word was flashed several times.

There was a practice period wherein there were 28 presentations of the seven words in Series E, and 8 presentations of the four words in Series D.

Following the warm-up period these same four words were presented in various randomly drawn series. After each eight exposures, the words were checked by calling them out, and *S*s were told to mark whether their responses were correct. Upon completion of two of these check series, the *S*s were moved backward or forward in the room until the entire group saw

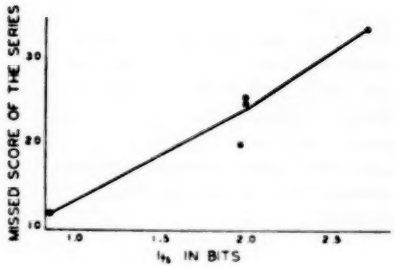

FIG. 1. Missed scores of each of the series plotted as a function of I_{ts}. The line passes through the mean point of the three conditions (B, C, D).

2 to 4 presentations correctly for two series. This procedure was introduced to equate as well as possible the differences in visual acuity of the *Ss*.

When the aforementioned criterion had been reached, the regular experimental procedure was begun. At this time the percentage figures with the words on the poster were explained. For example, *mean* had 25% beside it which indicated that it would appear ¼ of the time during the series.

The series were presented in the order A, B, C, D, E with the appropriate poster hung before each series and instructions given re-emphasizing the importance of writing a word only if the *S* was certain of its identity. A 2-min. rest period separated the successive series.

RESULTS

Figure 1 portrays the function obtained between the I_{ts} and the missed scores of the series. The function is apparently linear with the missed scores increasing as the I_{ts} increased.[2]

The effect of the I_{ts} upon missed scores was determined statistically with an analysis of variance (Table 2) (**5**). The obtained *F* ratios for the differences between groups of *Ss* and for the Groups × Series interaction fell short of significance at the .05 level. The *F* value for differences between the series was significant beyond the .001 level. These findings indicate that all *Ss* could be combined and the total performance to the series could be examined. The significant *F* value substantiates the hypothesis of an interaction effect.

[2] The obtained functions (Fig. 1 and 2) were not subjected to any formal test for nonlinearity because the author felt that inspection indicates no significant results.

TABLE 2

ANALYSIS OF VARIANCE OF THE "MISSED" SCORES

Source	*df*	*MS*	*F*
Between groups	2	1198	1.05
Between *Ss*	33	1139 (*e*)	
Between series	4	1984	16.09*
Groups × Series	8	149	1.21
Pooled *Ss* × Series	132	123 (*e*)	
Total	179		

Note:—(*e*) refers to the error term used to obtain the *F* scores above it.
* $P < .001$.

Tukey's "gaps" test (**2**, p. 251), based on the analysis of variance summarized in Table 2, indicated reliable differences ($P < .01$) between the mean of Series A and those of Series B, C, and D. Similarly, the mean of Series E differed from the others. No reliable differences could be demonstrated between the means of Series B, C, and D.

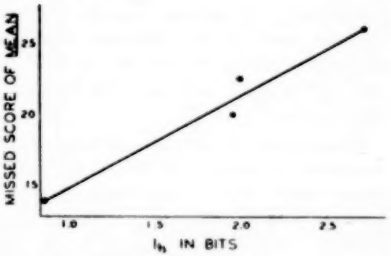

FIG. 2. Missed scores of *mean* plotted against I_{ts}. The I_{ss} value of *mean* was held constant at 2.0 bits. The effect of the I_{ts} upon the percentage of "missed" presentations of mean is given here. The line passes through the mean point of the two series at approximately 2.0 bits of I_{ts}.

TABLE 3

ANALYSIS OF VARIANCE OF THE MISSED SCORES FOR THE WORD "MEAN" WITH CONSTANT I_{s_s} IN FOUR SITUATIONS WITH DIFFERING I_{t_s}

Source	*df*	*MS*	F
Subjects	35	1432	9.65*
Series	3	912	6.14*
Ss × Series	105	148	
Total	143		

* $P = .001$.

By plotting the missed scores of *mean* against the I_{t_s} for each series there was obtained another function which was apparently linear (Fig. 2). Although the I_{s_s} remained constant at 2.00 bits, the missed scores increased linearly as the I_{t_s} increased.

The data were also tested for significance by an analysis of variance (**5**). As shown in Table 3, the variance between *S*s was significant. This, however, indicated individual differences in performance and was expected.

The significant between-series variance indicates that the linear function between I_{t_s} and I_{s_s} is composed of significantly differing performances to the constant information value of *mean* (2.00 bits). The interaction between I_{t_s}–I_{s_s} values (shown by this test) is that which effected the significant differences found in the *F* test of Table 2, and not the intra-I_{s_s} interaction.

Since the order of presentation remained constant—A, B, C, D, E—there was the possibility that a fatigue effect was confounding the results. This effect would likely be of the same nature as that predicted for the increasing I_{t_s} in this presentation order; therefore, an analysis was performed to reveal the fatigue effect. The structure of each series was examined as to the proportion of appearance of each word in the first half and the second half. The proportion of total misses for each was tabulated. There was no consistent trend from the first to the second half over the five series. This indicates that neither fatigue nor practice was a serious factor in the results.

DISCUSSION

The results obtained in this study substantiate the findings of Archer (**1**), Frick (**3**), and Hyman (**4**) that the I_{t_s} seems to affect behavior in a linear fashion: as I_{t_s} increases, ease of recognition decreases proportionately. Thus these findings indicate that the information measure for a total stimulus situation is a reliable method of characterizing responses to a group of visual stimuli.

The more important finding, however, was the significant interaction effect. The first type of interaction (among the stimuli themselves) was not present. This concurs with the results of Hyman (**4**) wherein he found that three different stimulus conditions with similar I_{t_s} values did not effect significantly different results. Thus it appears that differing configurations of stimuli within conditions of equal I_{t_s} do not interact among themselves to a significant degree.

The second type of interaction (I_{t_s}–I_{s_s}) was found. These results suggest that the entire stimulus condition plays the very important part of effecting what the reaction will be to a specific stimulus. This apparently occurs in a linear fashion. Therefore, the information measure *must* be used to characterize *both* the total and specific visual stimuli in order to obtain reliable predictions of responses to the I_{s_s} values.

SUMMARY

This study was designed to investigate the interaction (*a*) among stimuli within one condition and (*b*) between total stimulus information (I_{t_s}) and specific stimulus information (I_{s_s}) values. Thirty-six *S*s were presented 2-, 4-, 5-, 4- and 7-word series tachistoscopically in that order for a total of 40 exposures each.

The information load was varied for each series, and for each stimulus word in the series. The results were evaluated by a missed score measure. It was found that the I_{is} reliably characterized responses to total stimulus conditions; that is, the stimuli within a specific condition do not apparently interact. This substantiates the findings of previous authors. There was, however, a significant I_{is}–I_{ss} interaction. There is apparently a linear relationship between the responses to a constant I_{ss} value and varying I_{is} values.

References

1. Archer, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.
2. Edwards, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.
3. Frick, F. C. Some perceptual problems from the point of view of information theory. In *Current trends in information theory.* Pittsburgh: Univer. of Pittsburgh Press, 1953.
4. Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.
5. McNemar, Q. *Psychological statistics.* New York: Wiley, 1956.
6. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., 1944.

(Received April 19, 1957)

Journal of Experimental Psychology
Vol. 55, No. 3, 1958

SUPPLEMENTARY REPORT: EFFECT OF ADDITION OF IRRELEVANT VERBAL CUES ON PERCEPTUAL-MOTOR LEARNING[1]

HARRY W. BRAUN AND A. W. BENDIG

University of Pittsburgh

The authors have shown that the acquisition of a perceptual motor skill can be facilitated by the addition of irrelevant verbal cues when the cues were high- and low-meaningful terms selected from Noble's *m*-scale (**1**). The purpose of the present study was (*a*) to determine whether comparable facilitation of learning would occur when the irrelevant verbal cues were nonsense syllables of high and low association value, and (*b*) to test the prediction that as similarity among the irrelevant verbal cues increased, facilitation of learning would decrease.

Method.—The apparatus and problem sequences have been described (**1**). The 55 new *Ss* were drawn from the same population and divided into five equal groups. The control group learned a given 11-button sequence without verbal cues, as before. The *Ss* in the four experimental groups learned a sequence but also spelled aloud a nonsense syllable placed above each button: Group H-LS had syllables of high association value and low intralist similarity; Group H-HS had syllables of high association value and high intralist similarity; Group L-LS had syllables of low association value and low intralist similarity; and Group L-HS had syllables of low association value and high intralist similarity. Mean association values of the high- and low-meaningfulness lists were 93% and 16.8%, respectively, as determined by Glaze (**2**). Intralist similarity was varied in terms of frequency of repetition of first and last letters following a procedure described by Underwood (**3**).

Results.—The trials-to-criterion scores were converted by the logarithmic transformation used in the previous study (**1**). The present control group ($N = 11$) was combined with the previous control group ($N = 11$) because of similarity of means (68.2 and 67.6) and *SD*'s (27.5 and 27.0).

[1] The authors acknowledge the assistance of Lucille Mansfield and John Batrus in running the groups.

TABLE 1

MEANS AND *SD*'S OF TRIALS TO CRITERION

Group	*N*	Raw Scores		Log Scores	
		Mean	*SD*	Mean	*SD*
Control	22	67.9	27.2	1.79	.19
H-LS	11	40.6	10.2	1.60	.11
H-HS	11	56.3	20.3	1.73	.14
L-LS	11	47.7	7.4	1.67	.07
L-HS	11	46.0	12.1	1.65	.12

Scores of *S*s in the six groups (two control and four experimental) were subjected to an analysis of variance. The difference between the combined control groups and the combined experimental groups was significant ($F = 6.52$; $P = .05$) and Association Value × Similarity was also significant ($F = 5.64$; $P = .05$). However, neither Association Value alone nor Similarity alone was significant ($F = .00$ and 1.71). Since Control-Experimental differences and the Association Value-Similarity interaction were both significant, separate t tests were computed on the differences in mean log scores of the separate groups. The raw and log score means and *SD*'s of the groups are in Table 1.

The combined control groups were significantly inferior in learning performance ($P = .05$) when compared to Group H-LS ($t = 2.70$) and possibly ($P = .10$) when compared with Group L-HS ($t = 1.99$). The high similarity and low similarity groups are significantly different ($P = .05$) within the high association value condition ($t = 2.78$), but the two similarity groups are not different within the low association value condition ($t = .57$). In fact, the direction of differences between the means of the two similarity groups are reversed within the two levels of association value. None of the other t tests approached statistical significance.

Discussion.—The findings support the previous conclusion that the development of a perceptual-motor skill can be facilitated by the addition of irrelevant verbal cues (**1**). In both studies, the difference between the effects of high- and low-meaningfulness cues was not significant. The further prediction that facilitation of learning would decrease as similarity among the verbal cues increased was supported only in the case of nonsense syllables of high association value.

REFERENCES

1. BRAUN, H. W., & BENDIG, A. W. The effect of the addition of irrelevant verbal cues on perceptual motor learning. *J. exp. Psychol.*, 1957, **54**, 105–109.
2. GLAZE, J. A. The association value of non-sense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.
3. UNDERWOOD, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87.

(Received April 22, 1957)

ERRATUM

In the article "Intelligibility as a Function of Frequency of Usage" by M. R. Rosenzweig and L. Postman (*J. exp. Psychol.*, 1957, **54**, 412–422), the values in Table 2 (p. 414) for English subjects at the frequency range 100–330 should be the following: Mean, 5.18; *SD*, .59.

Publication Manual
of the American Psychological Association
1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. All scientists who are writing for publication will find the Publication Manual an indispensable guide.

Price $1.00

Discounts on quantity orders over 50 copies

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office, 1333 Sixteenth Street, N.W., Washington 6, D. C.

ERGONOMICS

Human Factors in Work, Machine Control and Equipment Design

General Editor

A. T. WELFORD

University of Cambridge, Psychological Laboratory, Downing Place, Cambridge

Editorial Board

Belgium: Dr. H. BASTENIER. Laboratoire d'Hygiène, Université Libre de Bruxelles.

France: Professor R. BONNARDEL. École Pratique des Hautes Études, Laboratoire de Psychologie Appliquée, 41 Rue Gay-Lussac, Paris 5.

Professor BERNARD METZ. Institut de Physiologie, Faculté de Médecine, Université de Strasbourg.

Germany: Professor E. A. MÜLLER. Max-Planck-Institut für Arbeitsphysiologie, Dortmund.

Great Britain: M. G. BENNETT. Research Department, British Railways, 20 Euston Square, London, N.W.1.

Dr. W. F. FLOYD. Department of Physiology, Middlesex Hospital Medical School, London, W.1.

Dr. W. E. HICK. University of Cambridge, Psychological Laboratory, Downing Place, Cambridge.

Emeritus Professor SIR CHARLES LOVATT EVANS. Hedgemoor Cottage, Winterslow, Near Salisbury.

Dr. L. G. NORMAN. London Transport Executive Medical Department, 280 Marylebone Road, London, N.W.1.

Netherlands: Dr. F. H. BONJER. Nederlands Instituut voor Praeventieve Geneeskunde, Leiden.

Sweden: Professor S. P. M. FORSSMAN. S. Blasieholmshamnen 4A, Stockholm, 16.

Switzerland: Professor E. GRANDJEAN. Institut für Hygiene und Arbeitsphysiologie der Eidgenössischen Technischen Hochschule, Zürich.

U.S.A.: Professor H. S. BELDING. University of Pittsburgh, Graduate School of Public Health, Pittsburgh 13, Pennsylvania.

Professor P. M. FITTS. Aviation Psychology Laboratory, Ohio State University, Columbus 10, Ohio.

Assistant Editor

Miss H. M. CLAY

Price £1 5s. 0d. per part plus postage

Subscription price per volume £4 15s. 0d. post free, payable in advance

Printed and Published by

TAYLOR AND FRANCIS LTD.

Red Lion Court, Fleet Street, London, E. C. 4.

Agents for U.S.A. and Canada

Academic Press Inc., 111 Fifth Avenue, New York 3, N.Y., U.S.A.

$3.50 per part. $13.30 per volume.